



10/500530
PCT/EP 02/14902

14 MAR 2003
01 JUL 2004



INVESTOR IN PEOPLE

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 27 MAR 2003

WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

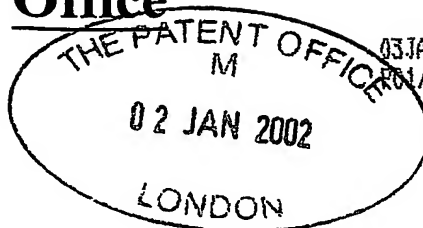
In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated

[Signature]
20 January 2003



03JAN02 E685083-1 C69803
01/7700 0.00-0200025.5

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

MJL/ND/B45292

2. Patent application number

(The Patent Office will fill in his part)

02 JAN 2002

0200025.5

3. Full name, address and postcode of the or of each applicant (underline all surnames)

GlaxoSmithKline Biologicals s.a.
Rue de l'Institut 89, B-1330 Rixensart, Belgium

Patents ADP number (if you know it) 08101271001

Belgian

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention

Novel Compounds

5. Name of your agent (if you have one)

Corporate Intellectual Property

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode)

GlaxoSmithKline
Two New Horizons Court
BRENTFORD
Middlesex TW8 9EP

Patents ADP number (if you know it)

07960982002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and (if you know it) the or each application number

Country	Priority application number (if you know it)	Date of filing (day / month / year)
---------	---	--

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing (day / month / year)
-------------------------------	--

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer yes if:

- a) any applicant named in part 3 is not an inventor, or
 - b) there is an inventor who is named as an applicant, or
 - c) any named applicant is a corporate body
- See note (d)

Novel Compounds

FIELD OF THE INVENTION

This invention relates to polynucleotides, (herein referred to as "BASB231 polynucleotide(s)"), polypeptides encoded by them (referred to herein as "BASB231" or "BASB231 polypeptide(s)"), recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including vaccines against bacterial infections. In a further aspect, the invention relates to diagnostic assays for detecting infection of certain pathogens.

BACKGROUND OF THE INVENTION

Haemophilus influenzae is a non-motile Gram negative bacterium. Man is its only natural host.

H. influenzae isolates are usually classified according to their polysaccharide capsule. Six different capsular types designated a through f have been identified. Isolates that fail to agglutinate with antisera raised against one of these six serotypes are classified as non typeable, and do not express a capsule.

The *H. influenzae* type b is clearly different from the other types in that it is a major cause of bacterial meningitis and systemic diseases. non typeable *H. influenzae* (NTHi) are only occasionally isolated from the blood of patients with systemic disease.

NTHi is a common cause of pneumonia, exacerbation of chronic bronchitis, sinusitis and otitis media.

Otitis media is an important childhood disease both by the number of cases and its potential sequelae. More than 3.5 millions cases are recorded every year in the United States, and it is estimated that 80 % of children have experienced at least one episode of otitis before reaching the age of 3 (1). Left untreated, or becoming chronic, this disease may lead to hearing loss that can be temporary (in the case of fluid accumulation in the

B45292

(17) is highly similar to the Hsf adhesin expressed by *H. influenzae* type b strains (18).

Another protein, the Hap protein shows similarity to IgA1 serine proteases and has been shown to be involved in both adhesion and cell entry (19).

- 5 Five major outer membrane proteins (OMP) have been identified and numerically numbered.

Original studies using *H. influenzae* type b strains showed that antibodies specific for P1 and P2 protected infant rats from subsequent challenge (20-21). P2 was found to be able
10 to induce bactericidal and opsonic antibodies, which are directed against the variable regions present within surface exposed loop structures of this integral OMP (22-23). The lipoprotein P4 also could induce bactericidal antibodies (24).

P6 is a conserved peptidoglycan-associated lipoprotein making up 1-5 % of the outer
15 membrane (25). Later a lipoprotein of about the same mol. wt. was recognized, called PCP (P6 crossreactive protein) (26). A mixture of the conserved lipoproteins P4, P6 and PCP did not reveal protection as measured in a chinchilla otitis-media model (27). P6 alone appears to induce protection in the chinchilla model (28).

20 P5 has sequence homology to the integral *Escherichia coli* OmpA (29-30). P5 appears to undergo antigenic drift during persistent infections with NTHi (31). However, conserved regions of this protein induced protection in the chinchilla model of otitis media.

25 In line with the observations made with gonococci and meningococci, NTHi expresses a dual human transferrin receptor composed of TbpA and TbpB when grown under iron limitation. Anti-TbpB protected infant rats. (32). Hemoglobin / haptoglobin receptors have also been described for NTHi (33). A receptor for Haem: Hemopexin has also been identified (34). A lactoferrin receptor is also present in NTHi, but is not yet characterized
30 (35).

18. St. Geme, JW. et al. (1996) *J. Bact.* 178:6281
19. St. Geme, JW. et al. (1994) *Mol. Microbiol.* 14:217
20. Loeb, MR. et al. (1987) *Infect. Immun.* 55:2612
21. Musson, RS. Jr. et al. (1983) *J. Clin. Invest.* 72:677
- 5 22. Haase, EM. et al. (1994) *Infect. Immun.* 62:3712
23. Troelstra, A. et al. (1994) *Infect. Immun.* 62:779
24. Green, BA. et al. (1991) *Infect. Immun.* 59:3191
25. Nelson, MB. et al. (1991) *Infect. Immun.* 59:2658
26. Deich, RM. et al. (1990) *Infect. Immun.* 58:3388
- 10 27. Green, BA. et al. (1993) *Infect. Immun.* 61:1950
28. Demaria, TF. et al. (1996) *Infect. Immun.* 64:5187
29. Miyamoto, N., Bakaletz, LO (1996) *Microb. Pathog.* 21:343
30. Munson, RS.jr. et al. (1993) *Infect. Immun.* 61:1017
31. Duim, B. et al. (1997) *Infect. Immun.* 65:1351
- 15 32. Loosmore, SM. et al (1996) *Mol. Microbiol.* 19:575
33. Maciver, I. et al. (1996) *Infect. Immun.* 64:3703
34. Cope, LD. et al. (1994) *Mol. Microbiol.* 13:868
35. Schryvers, AB. et al. (1989) *J. Med. Microbiol.* 29:121
36. Flack, FS. et al. (1995) *Gene* 156:97
- 20 37. Akkoyunlu, M. et al. (1996) *Infect. Immun.* 64:4586
38. Kimura, A. et al. (1985) *Infect. Immun.* 47:253
39. Mulks, MH. et Shoberg, RJ (1994) *Meth. Enzymol.* 235:543
40. Lomholt, H. Alphen, Lv, Kilian, M. (1993) *Infect. Immun.* 61:4575
41. Kyd, J.M. and Cripps, A.W. (1998) *Infect. Immun.* 66:2272
- 25 42. Loosmore, S.M. et al. (1998) *Infect. Immun.* 66:899

The frequency of NTHi infections has risen dramatically in the past few decades. This phenomenon has created an unmet medical need for new anti-microbial agents, vaccines, drug screening methods and diagnostic tests for this organism. The present invention

30 aims to meet that need.

					<i>influenzae</i> (74%)
Orf4	726	241	7	8	Molybdenum ABC transporter, periplasmic molybdate-binding protein, <i>Deinococcus radiodurans</i> (26%)
Orf5	741	246	9	10	ABC transporter, <i>Haemophilus influenzae</i> (38%)
Orf6	1023	340	11	12	ABC transporter, <i>Haemophilus influenzae</i> (45%)
Orf7	942	313	13	14	ABC transporter, <i>Haemophilus influenzae</i> (56%)
Orf8	558	185	15	16	Invasin precursor (YadA c-term), <i>Yersinia enterocolitica</i> (27%)
Orf9	2373	790	17	18	DNA methylase hsdm, <i>Vibrio cholerae</i> (70%)
Orf10	818	272	19	20	Leucyl tRNA synthetase, <i>Borrelia burgdorferi</i> (28%)
Orf11	636	211	21	22	ATP dependant DNA helicase, <i>Deinococcus radiodurans</i> (37%)
Orf12	1257	418	23	24	Type I restriction-modification system (s subunit), <i>Caulobacter crescentus</i> (29%)
Orf13	3027	1008	25	26	Type I restriction enzyme hsdR, <i>Vibrio cholerae</i> (65%)
Orf14	2052	683	27	28	Probable aaa family atpase, <i>Campylobacter jejuni</i> (33%)
Orf15	975	324	29	30	No homology with known protein
Orf16	744	247	31	32	Hypothetical 29.0 kd protein, <i>Aquifex aeolicus</i> (24%)
Orf17	846	271	33	34	Hypothetical 27.0 kd protein, <i>Aquifex aeolicus</i> (30%)
Orf18	273	90	35	36	Cell division protein ftsK (C-term), <i>Escherichia coli</i> (46%)
Orf19	1023	340	37	38	Putative dna-binding protein, <i>Neisseria meningitidis</i> (45%)
Orf20	711	236	39	40	Hypothetical 22.9 kd protein, <i>Actinobacillus actinomycetemcomitans</i> (79%)
Orf21	456	151	41	42	Yors protein, <i>Bacillus subtilis</i> (26%)
Orf22	441	146	43	44	Phosphate transport atp-binding protein pstB homolog, <i>Mycoplasma genitalium</i> (24%)
Orf23	642	213	45	46	No homology with known protein
Orf24	1344	447	47	48	Type I restriction protein, <i>Haemophilus influenzae</i> (40%)
Orf25	1995	664	49	50	Hypothetical 84.7 kda protein, <i>Thermotoga maritima</i> (25%)
Orf26	1155	384	51	52	Anticodon nuclease, <i>Neisseria meningitidis</i> (61%)
Orf27	999	332	53	54	wkue. gp8 protein, <i>wolbachia</i> sp. (40 %)
Orf28	819	272	55	56	Putative transposase protein, <i>Rhizobium meliloti</i> (40%)
Orf29	333	110	57	58	Partial sequence of <i>Bacteriophage</i> <i>ifl</i> . orf348 (35%)
Orf30	261	86	59	60	Putative cytoplasmic protein, <i>Salmonella typhimurium</i> lt2 (27%)
Orf31	927	308	61	62	Tryptophan 2-monooxygenase, <i>Agrobacterium</i>

membrane, or as purified LOS). In addition the enzyme may be isolated or recombinantly produced for its specific function to be used in vitro to produce novel synthetic oligosaccharide structures.

- 5 It is understood that sequences recited in the Sequence Listing below as "DNA" represent an exemplification of one embodiment of the invention, since those of ordinary skill will recognize that such sequences can be usefully employed in polynucleotides in general, including ribopolynucleotides.

The sequences of the BASB231 polynucleotides are set out in SEQ ID NO:1, 3, 5, 7, 9,
10 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73. SEQ Group 1 refers herein to any one of the polynucleotides set out in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71 or 73.

The sequences of the BASB231 encoded polypeptides are set out in SEQ ID NO:2, 4, 6, 8,
15 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72. SEQ Group 2 refers herein to any one of the encoded polypeptides set out in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 or 72.

20 Polypeptides

In one aspect of the invention there are provided polypeptides of non typeable *H. influenzae* referred to herein as "BASB231" and "BASB231 polypeptides" as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

25

The present invention further provides for:

- (a) an isolated polypeptide which comprises an amino acid sequence which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% or exact identity, to that of any sequence of SEQ Group 2;
- 30 (b) a polypeptide encoded by an isolated polynucleotide comprising a polynucleotide sequence which has at least 85% identity, preferably at least 90% identity, more preferably

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence selected from SEQ Group 2 or of variants thereof, such as a continuous series of residues that includes an amino- and/or carboxyl-terminal amino acid sequence. Degradation forms of the polypeptides of the invention produced by or in a host cell, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Further preferred fragments include an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from an amino acid sequence selected from SEQ Group 2 or an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids truncated or deleted from an amino acid sequence selected from SEQ Group 2.

Still further preferred fragments are those which comprise a B-cell or T-helper epitope, for example those fragments/peptides described in Example 10.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these fragments may be employed as intermediates for producing the full-length polypeptides of the invention.

Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

The polypeptides, or immunogenic fragments, of the invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in

Streptococcus pneumoniae which synthesize an N-acetyl-L-alanine amidase, amidase LytA, (coded by the *lytA* gene {Gene, 43 (1986) page 265-272}) an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LytA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E.coli* C-LytA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LytA fragment at its amino terminus has been described {Biotechnology: 10, (1992) page 795-798}. It is possible to use the repeat portion of the LytA molecule found in the C terminal end starting at residue 178, for example residues 188 - 305.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

It is most preferred that a polypeptide of the invention is derived from non typeable *H. influenzae*, however, it may preferably be obtained from other organisms of the same taxonomic genus. A polypeptide of the invention may also be obtained, for example, from organisms of the same taxonomic family or order.

Polynucleotides

B45292

strain 3224A cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a polynucleotide sequence given in SEQ Group 1, typically a library of clones of chromosomal DNA of non typeable *H. influenzae* strain 3224A in *E. coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent hybridization conditions. By sequencing the individual clones thus identified by hybridization with sequencing primers designed from the original polypeptide or polynucleotide sequence it is then possible to extend the polynucleotide sequence in both directions to determine a full length gene sequence. Conveniently, such sequencing is performed, for example, using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Direct genomic DNA sequencing may also be performed to obtain a full length gene sequence. Illustrative of the invention, each polynucleotide set out in SEQ Group 1 was discovered in a DNA library derived from non typeable *H. influenzae*.

Moreover, each DNA sequence set out in SEQ Group 1 contains an open reading frame encoding a protein having about the number of amino acid residues set forth in SEQ Group 2 with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known to those skilled in the art.

The polynucleotides of SEQ Group 1, between the start codon and the stop codon, encode respectively the polypeptides of SEQ Group 2. The nucleotide number of start codon and first nucleotide of stop codon are listed in table 2 for each polynucleotide of SEQ Group 1.

Table 2

Name	Start codon	1 st nucleotide of Stop codon
------	-------------	---

B45292

(a) a polynucleotide sequence which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% or exact identity, to any polynucleotide sequence from SEQ Group 1 over the entire length of the polynucleotide sequence from SEQ Group 1; or

(b) a polynucleotide sequence encoding a polypeptide which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% or 100% exact identity, to any amino acid sequence selected from SEQ Group 2, over the entire length of the amino acid sequence from SEQ Group 2.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than non typeable *H. influenzae*, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions (for example, using a temperature in the range of 45 – 65°C and an SDS

concentration from 0.1 – 1%) with a labeled or detectable probe consisting of or comprising any sequence selected from SEQ Group 1 or a fragment thereof; and isolating a full-length gene and/or genomic clones containing said polynucleotide sequence.

The invention provides a polynucleotide sequence identical over its entire length to a coding sequence (open reading frame) set out in SEQ Group 1. Also provided by the invention is a coding sequence for a mature polypeptide or a fragment thereof, by itself as well as a coding sequence for a mature polypeptide or a fragment in reading frame with another coding sequence, such as a sequence encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence. The polynucleotide of the invention may also contain at least one non-coding sequence, including for example, but not limited to at least one non-coding 5' and 3' sequence, such as the transcribed but non-translated sequences, termination signals (such as rho-dependent and rho-independent termination signals), ribosome binding sites, Kozak sequences, sequences that stabilize mRNA, introns, and polyadenylation signals.

The polynucleotide sequence may also comprise additional coding sequence encoding additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker

Orf25	1	1992
Orf26	1*	1152
Orf27	1	996
Orf28	1	816
Orf29	1*	330
Orf30	1	258
Orf31	1	915
Orf32	1*	309
Orf33	1	1461
Orf34	1	885
Orf35	1*	840
Orf36	1*	390
Orf37	1	672

*It is not the start codon but it is the first nucleotide of the coding sequence

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the non typeable *H. influenzae* BASB231 having an amino acid sequence set out in any of the sequences of SEQ Group 2 .

The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, polynucleotides interrupted by integrated phage, an integrated insertion sequence, an integrated vector sequence, an integrated transposon sequence, or due to RNA editing or genomic DNA reorganization) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode variants of a polypeptide having a deduced amino acid sequence of any of the sequences of SEQ Group 2 . Fragments of polynucleotides of the invention may be used, for example, to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding BASB231 variants, that have the amino acid sequence of BASB231 polypeptide of any sequence from SEQ Group 2 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, modified, deleted and/or added, in any combination. Especially preferred

solution, 10% dextran sulfate, and 20 micrograms/ml of denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C.

Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y.,

5 (1989), particularly Chapter 11 therein. Solution hybridization may also be used with the polynucleotide sequences provided by the invention.

The invention also provides a polynucleotide consisting of or comprising a polynucleotide sequence obtained by screening an appropriate library containing the complete gene for a
10 polynucleotide sequence set forth in any of the sequences of SEQ Group 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in the corresponding sequence of SEQ Group 1 or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers fully described elsewhere herein.

15

As discussed elsewhere herein regarding polynucleotide assays of the invention, for instance, the polynucleotides of the invention, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding BASB231 and to isolate cDNA and genomic clones of other genes that have a high identity,
20 particularly high sequence identity, to the BASB231 gene. Such probes generally will comprise at least 15 nucleotide residues or base pairs. Preferably, such probes will have at least 30 nucleotide residues or base pairs and may have at least 50 nucleotide residues or base pairs. Particularly preferred probes will have at least 20 nucleotide residues or base pairs and will have less than 30 nucleotide residues or base pairs.

25

A coding region of a BASB231 gene may be isolated by screening using a DNA sequence provided in SEQ Group 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the
30 probe hybridizes to.

instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

For each and every polynucleotide of the invention there is provided a polynucleotide complementary to it. It is preferred that these complementary polynucleotides are fully complementary to each polynucleotide with which they are complementary.

A precursor protein, having a mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleotides, the term "N" may also be used in describing certain polynucleotides of the invention. "N" means that any of the four DNA or RNA nucleotides may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a nucleic acid that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis, *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook, *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, conjugation, transduction, scrape loading, ballistic introduction and infection.

- 10 Representative examples of appropriate hosts include bacterial cells, such as cells of streptococci, staphylococci, enterococci, *E. coli*, streptomyces, cyanobacteria, *Bacillus subtilis*, *Neisseria meningitidis*, *Haemophilus influenzae* and *Moraxella catarrhalis*; fungal cells, such as cells of a yeast, *Kluveromyces*, *Saccharomyces*, *Pichia*, a basidiomycete, *Candida albicans* and *Aspergillus*; insect cells such as cells of *Drosophila* S2 and
- 15 *Spodoptera* Sf9; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293, CV-1 and Bowes melanoma cells; and plant cells, such as cells of a gymnosperm or angiosperm.

- A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal-, episomal- and virus-derived
- 20 vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, picornaviruses, retroviruses, and alphaviruses and vectors derived from combinations thereof, such as those derived from
- 25 plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any
- 30 of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

diagnostic method for diagnosis of disease, staging of disease or response of an infectious organism to drugs. Eukaryotes, particularly mammals, and especially humans, particularly those infected or suspected to be infected with an organism comprising the BASB231 gene or protein, may be detected at the nucleic acid or amino acid level by a variety of well known techniques as well as by methods provided herein.

Polypeptides and polynucleotides for prognosis, diagnosis or other analysis may be obtained from a putatively infected and/or infected individual's bodily materials. Polynucleotides from any of these sources, particularly DNA or RNA, may be used directly for detection or may be amplified enzymatically by using PCR or any other amplification technique prior to analysis. RNA, particularly mRNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of infectious or resident organism present in an individual, may be made by an analysis of the genotype of a selected polynucleotide of the organism. Deletions and insertions can be detected by a change in size of the amplified product in comparison to a genotype of a reference sequence selected from a related organism, preferably a different species of the same genus or a different strain of the same species. Point mutations can be identified by hybridizing amplified DNA to labeled BASB231 polynucleotide sequences. Perfectly or significantly matched sequences can be distinguished from imperfectly or more significantly mismatched duplexes by DNase or RNase digestion, for DNA or RNA respectively, or by detecting differences in melting temperatures or renaturation kinetics. Polynucleotide sequence differences may also be detected by alterations in the electrophoretic mobility of polynucleotide fragments in gels as compared to a reference sequence. This may be carried out with or without denaturing agents. Polynucleotide differences may also be detected by direct DNA or RNA sequencing. See, for example, Myers *et al.*, *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase, V1 and S1 protection assay or a chemical cleavage method. See, for example, Cotton *et al.*, *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).

In another embodiment, an array of oligonucleotides probes comprising BASB231 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening

example. For example, RT-PCR can be used to detect mutations in the RNA. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA, cDNA or genomic DNA may also be used for the same purpose, PCR. As an example, PCR primers complementary to a polynucleotide encoding BASB231 polypeptide can be used to identify and analyze mutations.

The invention further provides primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among other things, amplifying BASB231 DNA and/or RNA isolated from a sample derived from an individual, such as a bodily material. The primers may be used to amplify a polynucleotide isolated from an infected individual, such that the polynucleotide may then be subject to various techniques for elucidation of the polynucleotide sequence. In this way, mutations in the polynucleotide sequence may be detected and used to diagnose and/or prognose the infection or its stage or course, or to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections caused by non typeable *H. influenzae*, comprising determining from a sample derived from an individual, such as a bodily material, an increased level of expression of polynucleotide having a sequence of any of the sequences of SEQ Group 1. Increased or decreased expression of BASB231 polynucleotide can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting, spectrometry and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of BASB231 polypeptide compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of BASB231 polypeptide, in a sample derived from a host, such as a bodily material, are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis, antibody sandwich assays, antibody detection and ELISA assays.

preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides or polynucleotides of this invention. Also, transgenic mice, or other organisms or animals, such as other mammals, may be used to express humanized antibodies immunospecific to the polypeptides or polynucleotides of the invention.

Alternatively, phage display technology may be utilized to select antibody genes with binding activities towards a polypeptide of the invention either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-BASB231 or from naive libraries (McCafferty, *et al.*, (1990), *Nature* 348, 552-554; Marks, *et al.*, (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by, for example, chain shuffling (Clackson *et al.*, (1991) *Nature* 352: 628).

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides or polynucleotides of the invention to purify the polypeptides or polynucleotides by, for example, affinity chromatography.

Thus, among others, antibodies against BASB231 polypeptide or BASB231 polynucleotide may be employed to treat infections, particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants form a particular aspect of this invention.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be

standard. Fusion proteins, such as those made from Fc portion and BASB231 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists of the polypeptide of the present invention, as well as of phylogenetically and and/or functionally related polypeptides (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies that bind to and/or interact with a polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and/or polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of BASB231 polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bactericidal.

The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising BASB231 polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a BASB231 agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the BASB231 polypeptide is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of BASB231 polypeptide are most likely to be good antagonists. Molecules that bind well and, as the case may be, increase the rate of product production from substrate, increase signal transduction, or increase chemical channel activity are agonists. Detection of the rate or level of, as the case may be, production of product from substrate, signal transduction, or

B45292

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Each of the polynucleotide sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the polynucleotide sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide, agonist or antagonist of the invention to interfere with the initial physical interaction between a pathogen or pathogens and a eukaryotic, preferably mammalian, host responsible for sequelae of infection. In particular, the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive and/or gram negative bacteria, to eukaryotic, preferably mammalian, extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block bacterial adhesion between eukaryotic, preferably mammalian, extracellular matrix proteins and bacterial BASB231 proteins that mediate tissue damage and/or; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

capable of binding to anti-native peptide antibodies, but may not necessarily themselves share significant sequence homology to the native polypeptide.

Vaccines

- 5 Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal, preferably humans, which comprises inoculating the individual with BASB231 polynucleotide and/or polypeptide, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most
- 10 particularly non typeable *H. influenzae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector, sequence or ribozyme to direct expression of BASB231 polynucleotide and/or polypeptide, or a fragment or a
- 15 variant thereof, for expressing BASB231 polynucleotide and/or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual, preferably a human, from disease, whether that disease is already established within the individual or not. One
- 20 example of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a ribozyme, a modified nucleic acid, a DNA/RNA hybrid, a DNA-protein complex or an RNA-protein complex.
- 25 A further aspect of the invention relates to an immunological composition that when introduced into an individual, preferably a human, capable of having induced within it an immunological response, induces an immunological response in such individual to a BASB231 polynucleotide and/or polypeptide encoded therefrom, wherein the composition comprises a recombinant BASB231 polynucleotide and/or polypeptide encoded therefrom
- 30 and/or comprises DNA and/or RNA which encodes and expresses an antigen of said BASB231 polynucleotide, polypeptide encoded therefrom, or other polypeptide of the

B45292

Blebs have the advantage of providing outer-membrane proteins in their native conformation and are thus particularly useful for vaccines. Blebs can also be improved for vaccine use by engineering the bacterium so as to modify the expression of one or more molecules at the outer membrane. Thus for example the expression of a desired immunogenic protein at the outer membrane, such as the BASB231 polypeptide, can be introduced or upregulated (e.g. by altering the promoter). Instead or in addition, the expression of outer-membrane molecules which are either not relevant (e.g. unprotective antigens or immunodominant but variable proteins) or detrimental (e.g. toxic molecules such as LPS, or potential inducers of an autoimmune response) can be downregulated. These approaches are discussed in more detail below.

The non-coding flanking regions of the BASB231 gene contain regulatory elements important in the expression of the gene. This regulation takes place both at the transcriptional and translational level. The sequence of these regions, either upstream or downstream of the open reading frame of the gene, can be obtained by DNA sequencing. This sequence information allows the determination of potential regulatory motifs such as the different promoter elements, terminator sequences, inducible sequence elements, repressors, elements responsible for phase variation, the shine-dalgarno sequence, regions with potential secondary structure involved in regulation, as well as other types of regulatory motifs or sequences. This sequence is a further aspect of the invention.

Furthermore, SEQ ID NO: 75 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs1, 2, 3, 4, 5, 6, 7, 8 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 75. The localisation of the ORFs of SED ID NO: 75 are listed in table 4.

Table 4:

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf1	90	542	+

B45292

Orf20	7027	6317	-
Orf21	7467	7011	-
Orf22	7966*	7526	-

*It is not the first nucleotide of the start codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 78 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 23, 24 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 78. The localisation of the ORFs of SED ID NO: 78 are listed in table 7.

Table 7

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf23	688	47	-
Orf24	2028	685	-

Furthermore, SEQ ID NO: 79 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 25 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 79. The localisation of the ORF of SED ID NO: 79 are listed in table 8.

Table 8

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf25	2205	211	-

Furthermore, SEQ ID NO: 80 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 26, 27 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 80. The localisation of the ORFs of SED ID NO: 80 are listed in table 9.

Table 9

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
------	---	---	--------

Table 12

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf33	74	1537	+

Furthermore, SEQ ID NO: 84 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 34 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 84. The localisation of the ORF of SED ID NO: 84 are listed in table 13.

Table 13

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf34	82	969	+

Furthermore, SEQ ID NO: 85 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 35 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 83. The localisation of the ORF of SED ID NO: 85 are listed in table 13.

Table 13

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf35	1065*	223	-

*It is not the first nucleotide of the start codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 86 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 36 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 86. The localisation of the ORF of SED ID NO: 86 are listed in table 14.

Table 14

Name	Position of the first nucleotide of	Position of the last nucleotide of stop	Strand
------	-------------------------------------	---	--------

on gene expression can be assessed. In another approach, the sequence knowledge of the region of interest can be used to replace or delete all or part of the natural regulatory sequences. In this case, the regulatory region targeted is isolated and modified so as to contain the regulatory elements from another gene, a combination of regulatory elements from different genes, a synthetic regulatory region, or any other regulatory region, or to delete selected parts of the wild-type regulatory sequences. These modified sequences can then be reintroduced into the bacterium via homologous recombination into the genome. A non-exhaustive list of preferred promoters that could be used for up-regulation of gene expression includes the promoters *porA*, *porB*, *lbpB*, *tbpB*, *p110*, *lst*, *hpuAB* from *N. meningitidis* or *N. gonorrhoeae*; *ompCD*, *copB*, *lbpB*, *ompE*, *UspA1*; *UspA2*; *TbpB* from *M. Catarrhalis*; *p1*, *p2*, *p4*, *p5*, *p6*, *lpD*, *tbpB*, *D15*, *Hia*, *Hmw1*, *Hmw2* from *H. influenzae*.

In one example, the expression of the gene can be modulated by exchanging its promoter with a stronger promoter (through isolating the upstream sequence of the gene, in vitro modification of this sequence, and reintroduction into the genome by homologous recombination). Upregulated expression can be obtained in both the bacterium as well as in the outer membrane vesicles shed (or made) from the bacterium.

In other examples, the described approaches can be used to generate recombinant bacterial strains with improved characteristics for vaccine applications. These can be, but are not limited to, attenuated strains, strains with increased expression of selected antigens, strains with knock-outs (or decreased expression) of genes interfering with the immune response, strains with modulated expression of immunodominant proteins, strains with modulated shedding of outer-membrane vesicles.

Thus, also provided by the invention is a modified upstream region of the BASB231 gene, which modified upstream region contains a heterologous regulatory element which alters the expression level of the BASB231 protein located at the outer membrane. The upstream region according to this aspect of the invention includes the sequence upstream of the BASB231 gene. The upstream region starts immediately upstream of the BASB231

Such experiments will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value, derived from the requisite organ of the animal successfully resisting or clearing infection, for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly non typeable *H. influenzae* infection, in mammals, particularly humans.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant polypeptide and/or polynucleotide of the invention together with a suitable carrier, such as a pharmaceutically acceptable carrier. Since the polypeptides and polynucleotides may be broken down in the stomach, each is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostatic compounds and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

The vaccine formulation of the invention may also include adjuvant systems for enhancing the immunogenicity of the formulation. Preferably the adjuvant system raises preferentially a TH1 type of response.

An immune response may be broadly distinguished into two extreme categories, being a humoral or cell mediated immune responses (traditionally characterised by antibody and cellular effector mechanisms of protection respectively). These categories of response have been termed TH1-type responses (cell-mediated response), and TH2-type immune responses (humoral response).

Thus, a TH1-type adjuvant is one which preferentially stimulates isolated T-cell populations to produce high levels of TH1-type cytokines when re-stimulated with antigen *in vitro*, and promotes development of both CD8+ cytotoxic T lymphocytes and antigen specific immunoglobulin responses associated with TH1-type isotype.

Adjuvants which are capable of preferential stimulation of the TH1 cell response are described in International Patent Application No. WO 94/00153 and WO 95/17209.

3 De-O-acylated monophosphoryl lipid A (3D-MPL) is one such adjuvant. This is known from GB 2220211 (Ribi). Chemically it is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains and is manufactured by Ribi Immunochem, Montana. A preferred form of 3 De-O-acylated monophosphoryl lipid A is disclosed in European Patent 0 689 454 B1 (SmithKline Beecham Biologicals SA).

Preferably, the particles of 3D-MPL are small enough to be sterile filtered through a 0.22micron membrane (European Patent number 0 689 454).

3D-MPL will be present in the range of 10µg - 100µg preferably 25-50µg per dose wherein the antigen will typically be present in a range 2-50µg per dose.

Another preferred adjuvant comprises QS21, an Hplc purified non-toxic fraction derived from the bark of *Quillaja Saponaria Molina*. Optionally this may be admixed with 3 De-O-acylated monophosphoryl lipid A (3D-MPL), optionally together with an carrier.

The method of production of QS21 is disclosed in US patent No. 5,057,540.

Non-reactogenic adjuvant formulations containing QS21 have been described previously (WO 96/33739). Such formulations comprising QS21 and cholesterol have been shown to be successful TH1 stimulating adjuvants when formulated together with an antigen.

Non-toxic oil in water emulsions preferably contain a non-toxic oil, e.g. squalane or squalene, an emulsifier, e.g. Tween 80, in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline.

- 5 A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil in water emulsion is described in WO 95/17210.

While the invention has been described with reference to certain BASB231 polypeptides and polynucleotides, it is to be understood that this covers fragments of the naturally occurring polypeptides and polynucleotides, and similar polypeptides and polynucleotides with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant polypeptides or polynucleotides.

10 The present invention also provides a polyvalent vaccine composition comprising a vaccine formulation of the invention in combination with other antigens, in particular antigens useful for treating *otitis media*. Such a polyvalent vaccine composition may include a TH-1 inducing adjuvant as hereinbefore described.

20 In a preferred embodiment, the polypeptides, fragments and immunogens of the invention are formulated with one or more plain or conjugated pneumococcal capsular polysaccharides, and one or more antigens that can protect a host against *M. catarrhalis* infection. Optionally, the vaccine may also comprise one or more protein antigens that can protect a host against *Streptococcus pneumoniae* infection. Optionally, the vaccine may also comprise one or more further non typeable *Haemophilus influenzae* protein antigens.

25 The vaccine may also optionally comprise one or more antigens that can protect a host against RSV and/or one or more antigens that can protect a host against influenza virus. Such a vaccine may be advantageously used as a global otitis media vaccine.

30 The pneumococcal capsular polysaccharide antigens are preferably selected from serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F,

5843464 - Ohio State Research Foundation), OMP26, P6, protein D, TbpA, TbpB, Hia, Hmw1, Hmw2, Hap, and D15.

Preferred influenza virus antigens include whole, live or inactivated virus, split
5 influenza virus, grown in eggs or MDCK cells, or Vero cells or whole flu virosomes (as
described by R. Gluck, Vaccine, 1992, 10, 915-920) or purified or recombinant proteins
thereof, such as HA, NP, NA, or M proteins, or combinations thereof.

Preferred RSV (Respiratory Syncytial Virus) antigens include the F glycoprotein, the G
10 glycoprotein, the HN protein, or derivatives thereof.

Compositions, kits and administration

In a further aspect of the invention there are provided compositions comprising a BASB231
15 polynucleotide and/or a BASB231 polypeptide for administration to a cell or to a
multicellular organism.

The invention also relates to compositions comprising a polynucleotide and/or a
polypeptides discussed herein or their agonists or antagonists. The polypeptides and
20 polynucleotides of the invention may be employed in combination with a non-sterile or
sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical
carrier suitable for administration to an individual. Such compositions comprise, for
instance, a media additive or a therapeutically effective amount of a polypeptide and/or
polynucleotide of the invention and a pharmaceutically acceptable carrier or excipient. Such
25 carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol,
ethanol and combinations thereof. The formulation should suit the mode of administration.
The invention further relates to diagnostic and pharmaceutical packs and kits comprising
one or more containers filled with one or more of the ingredients of the aforementioned
compositions of the invention.

30

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Sequence Databases, Sequences in a Tangible Medium, and Algorithms

Polynucleotide and polypeptide sequences form a valuable information resource with which to determine their 2- and 3-dimensional structures as well as to identify further sequences of similar homology. These approaches are most easily facilitated by storing the sequence in a

DEFINITIONS

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heine, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GAP program in the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990), and FASTA(Pearson and Lipman *Proc. Natl. Acad. Sci. USA* 85: 2444-2448 (1988). The BLAST family of programs is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Henikoff and Henikoff,

Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 8

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n . Alterations of polynucleotide sequences encoding the polypeptides of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequences of SEQ ID NO:1, that is it may be 100% identical, or it may include up to a certain integer number of nucleic acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of nucleic acids in SEQ ID NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleic acids in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleic acid alterations, x_n is the total number of nucleic acids in SEQ ID NO:1, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n .

reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

"Individual(s)," when used herein with reference to an organism, means a multicellular eukaryote, including, but not limited to a metazoan, a mammal, an ovid, a bovid, a simian, a primate, and a human.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA including single and double-stranded regions.

EXAMPLES:

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1: Cloning of the BASB231 gene from non typeable *Haemophilus influenzae* strain 3224A.

Genomic DNA is extracted from the non typeable *Haemophilus influenzae* strain 3224A from 10¹⁰ bacterial cells using the QIAGEN genomic DNA extraction kit (Qiagen GmbH). This material (1µg) is then submitted to Polymerase Chain Reaction DNA amplification using two specific primers. A DNA fragment is obtained, digested by the suitable restriction endonucleases and inserted into the compatible sites of the pET cloning/expression vector (Novagen) using standard molecular biology techniques (Molecular Cloning, a Laboratory Manual, Second Edition, Eds: Sambrook, Fritsch & Maniatis, Cold Spring Harbor press 1989). Recombinant pET-BASB231 is then submitted to DNA sequencing using the Big Dyes kit (Applied biosystems) and analyzed on a ABI 373/A DNA sequencer in the conditions described by the supplier.

Example 2: Expression and purification of recombinant BASB231 protein in *Escherichia coli*.

The construction of the pET-BASB231 cloning/expression vector is described in Example

1. This vector harbours the BASB231 gene isolated from the non typeable *Haemophilus influenzae* strain 3224A in fusion with a stretch of 6 Histidine residues, placed under the control of the strong bacteriophage T7 gene 10 promoter. For expression study, this vector is introduced into the *Escherichia coli* strain Novablue (DE3) (Novagen), in which, the gene for the T7 polymerase is placed under the control of the isopropyl-beta-D thiogalactoside (IPTG)-regulatable *lac* promoter. Liquid cultures (100 ml) of the Novablue (DE3) [pET-BASB231] *E. coli* recombinant strain are grown at 37°C under

B45292

calculated by 4-parameter logistic model using the XL Fit software. The antisera are also used as the first antibody to identify the protein in a western blot as described in example 5 below.

5 **Example 4: Immunological characterization: Surface exposure of BASB231**

Anti-BASB231 protein titres are determined by an ELISA using formalin-killed whole cells of non typable *Haemophilus influenzae* (NTHi). The titer is defined as mid-point titers calculated by 4-parameter logistic model using the XL Fit software.

10 **Example 5. Immunological Characterisation: Western Blot Analysis**

Several strains of NTHi, as well as clinical isolates, are grown on Chocolate agar plates for 24 hours at 36°C and 5% CO₂. Several colonies are used to inoculate Brain Heart Infusion (BHI) broth supplemented by NAD and hemin, each at 10 µg/ml. Cultures are grown until the absorbance at 620nm is approximately 0.4 and cells are collected by

15 centrifugation. Cells are then concentrated and solubilized in PAGE sample buffer. The solubilized cells are then resolved on 4-20% polyacrylamide gels and the separated proteins are electrophoretically transferred to PVDF membranes. The PVDF membranes are then pretreated with saturation buffer. All subsequent incubations are carried out using this pretreatment buffer.

20

PVDF membranes are incubated with preimmune serum or rabbit or mouse immune serum. PVDF membranes are then washed.

PVDF membranes are incubated with biotin-labeled sheep anti-rabbit or mouse Ig.

PVDF membranes are then washed 3 times with wash buffer, and incubated with

25

streptavidin-peroxydase. PVDF membranes are then washed 3 times with wash buffer and developed with 4-chloro-1-naphtol.

Example 6: Immunological characterization: Bactericidal Activity

Complement-mediated cytotoxic activity of anti-BASB231 antibodies is examined to

30

determine the vaccine potential of BASB231 protein antiserum that is prepared as

B45292

Mice are killed between 30 minutes and 24 hours after challenge and the lungs are removed aseptically and homogenized individually. The log₁₀ weighted mean number of CFU/lung is determined by counting the colonies grown on agar plates after plating of dilutions of the homogenate. The arithmetic mean of the log₁₀ weighted mean

5 number of CFU/lung and the standard deviations are calculated for each group.

• Results are analysed statistically.

In this experiment groups of mice are immunized either with BASB231 or with a killed whole cells (kwc) preparation of NTHi or sham immunized.

10

Example 9: Inhibition of NTHi adhesion onto cells by anti-BASB231 antiserum.

This assay measures the capacity of anti BASB231 sera to inhibit the adhesion of NTHi bacteria to epithelial cells. This activity could prevent colonization of the nasopharynx by NTHi.

15 One volume of bacteria is incubated on ice with one volume of pre-immune or anti-BASB231 immune serum dilution. This mixture is subsequently added in the wells of a 24 well plate containing a confluent cells culture that is washed once with culture medium to remove traces of antibiotic. The plate is centrifuged and incubated.

Each well is then gently washed. After the last wash, sodium glycocholate is added to
20 the wells. After incubation, the cell layer is scraped and homogenised. Dilutions of the homogenate are plated on agar plates and incubated. The number of colonies on each plate is counted and the number of bacteria present in each well calculated.

25

SEQUENCE INFORMATION**BASB231 Polynucleotide and Polypeptide Sequences****5 SEQ ID NO:1 polynucleotide sequence of Orf1**

GTGTGCTATGAGCCATTTATTTATTACCCAATGATGTGCAATGAAAAGATAGCGCGTGCTATTATTCTTG
AAGATGATGCGATTGTATCGCACGAATTCGAAGCAATTGTAAAAGACAGTTTGAAGAAAGTTTCAAAAAA
TGTTGAAATTTTATTTTATGATCATGGTAAAGCAAAAAGTTATTGCTGGAAAAAACACTTGTCAAAAAT
TACCGTTTAGTTCACTATCGTAAACCCTCTAAAACGTCTAAACGTGCAATCATGTGTACAACAGCTTATT
10 TAATTACTTTTATCTGGCGCTCAAAAACCTCTACAAATAGCCTATCCTATCCGTATGCCTGCTGACTACTT
AACTGTGCTTTACAATTAAGTGGACTAAAGGCTTATGGTGTGAACCACCTTGTGTATTTAAAGGCGCA
ATTTCAAGAAATTGATGCAATGGAGCAACGCTAA

SEQ ID NO:2 polypeptide sequence of Orf1

VCYEPFIYYPMMCNEKIARAIILEDDAIVSHEFEAIVKDSLKKVSKNVEILFYDHGKAKSYCWKKTIVKNYR
15 LVHYRKPSKTSKRAIMCTTAYLITLSGAQKLLQIAYPIRMPADYLTGALQLTGLKAYGVEPPCVFKGAISEI
DAMEQR.

SEQ ID NO:3 polynucleotide sequence of Orf2

ATGAAATTAAAAATAAATTACAAATGTTAAGGTTGGGTCTAGGCAAATATTTCTTGATAAAAAAACG
GATTAAACAGAATAACAAATGTTCTAGAGCATCCTCTTCTCCGCAAGACGGAAAAATTGGGGATTA
20 TGTGGTGAGCTCATTTGTATTCGTGAGATAAAAAATTTAATCCCCACATTAAAATTGGTGTAAATTTGT
ACCAACAAAAATGCTTATCTTTTAAACAAAATCCATATATCGATCAACTTTACTATGTAAAAAGAAAA
GTATTTTGGATTACATCAAATGTGGTCTAGCAATTCAAAAAGAACAAATATGATTTAGTGATTGATCCGAC
GATTATGATTCGTAATCGCGATCTTTTACTTTTACGCTTAATCAATGCCAAGCATTATATTGGCTACCAA
AAAGCCAATTATGGTTTATTTAATATTAATCTGGAGGGACAATTTCACTTTTCGGAACCTCTATAAACTCG
25 CCTAGAAAAAGTGAATATTACGGTACAAGATATAAGCTATGACATCCCATTTGATAAGCAAAGTGCGGT
CGAAATTTCTGAATTTTTCAGAAAAACCACTAGAAAAAGTATATTGCTATTAATTTTATGGTGTCTGCA
AGAATCAAAAAGTAAACAATGACAACATCAAAAATATTTAGATTATCTCAGCAAGTCCGCGGAGGAA
AAAAGCTGGTGTATTAAAGCTATCCTGAAGTAACAGAGAAATTAACACAATTGTGAGCCGATTATCCGCA
TATTTTGTCCATCCAACAACCAAGATCTTTCATACCATTGAATTGATTGCGCACTGTGATCAATTAATC
30 TCTACAGACACGTCTACTGTACATATTGCTTCAGGTTTTAATAAACCAATTATTGGTATTTATAAAGAAG
ATCCTATTGCGTTTACACATTGGCAACCCAGAAGTCGGGCAGAAACGCACATACTTTTCTATAAAGAAAA
TATTAATGAGCTCTACCTGAACAAATTGACCCTGCATGGCTTGTCAAATAG

SEQ ID NO:4 polypeptide sequence of Orf2

MKLKNKLQMLRLGLGKYFLDKKNGLNRIITNVRPILFLRQDGKIGDYVVSSVFVREIKKFNPHIKIGVICTK
35 QNAYLQKQNPYIDQLYYVKKKSILDYIKCGLAIQKEQYDLVIDPTIMIRNRDLLLLRLINAKHYIGYQKANY
GLFNINLEGQFHFSELYKLALAKVNITVQDISYDIPFDKQSAVEISEFLQKNQLEKYIAINFYGAARIKKVN
NDNIKKYLDYLTQVRGGKLVLLSYPEVTEKLTQLSADYPHIFVHPTTKIFHTIELIRHCDQLISTDTSTVH
IASGFNKPILGIYKEDPIAFTHWQPRSRATHILFYKENINELSPQIDPAWLK.

SEQ ID NO:5 polynucleotide sequence of Orf3

ATGCCAGAATTACCTGAAGTTGAAACCACAAAAAATGGAATTAGCCCTTATCTTGAAGGGGCTATCATTG
40 AAAAAATTGTTGTTTCGCCAACCGAAATTACGCTGGATGGTAAGCGAAGAATTAGCGCAAATTACACAACA

B45292

MNELSLDADKLLFGYDKPLYLPLTFQCKKGEVISVFGTNGKGTLLHSLAHVLPVMSGQIRQQGHIGFVPQ
SFSSPDYPVLEIVLMGRASKIGAFNLPSKTDETVLQMLACLDILHLAERNINMLSGGQRQLVLIARALATE
CQVLILDEPTAALDVYNQXRVLQLIRFLATEQKMTIIFSTHDPYHSLCVADNVLLLLLPNQWKYGIASQILT
ESHLKQAYNVPIKYSMIEEQQVLVPIFTIQ.

5 **SEQ ID NO:11 polynucleotide sequence of Orf6**

ATGAAGTCTATGTTAGCAAATCAGCGAGGTTTTATAACATCGCTGATTTTTATCTTGTATTATCATCGTAT
TGTTCACTTTAAATATTGGCACTTTTTCGTTATCAACCGGAAAAGTGATGTCCATTTTATCTAAGCCTTT
TCTTTTCGCAACACGCGTCTTTTACACCTATGGAATACCATATTGTTTGGCATGTACGCTTACCACGCATC
ATTATGGCATTTTTTTTTTCAGGGGGGATCTGAGCGATGAGTGGTGCAACACTACAGGGCGTTTTTCATAATC
10 CCCTTGTTGATCCTCATATTATTGGTGTACATCAGGGGCAGTTTTTGGAGGCAGTTTAGCAATTTTATT
AGGATTCCCATCTTATTTATTGATTCTATCCACATTTTCTTTTGGTTTATTGACATTATTCTTGATCTAT
GTAACCACAATGTTTCATCGGAAAAGGCAATCGTATTGTATTAGTTTTAGCGGGTGTCAATTTAAGTGGTT
TCTTTAGCACTCTAGTGAGCTTAATCCAATATTTAGCGGATGCAGAAGAAGTTCTGCCGAGCATTGTATT
TTGGTTATTAGGAAGTTTTGCCACCACTAGTTGGGCAAACTAGCTATATTGTTACCCTGCGTTTTTATT
15 GCAGCTTATTTATTATTCCGTTTACGGTGGCATATTAATGTGTTATCGCTAGGTGATATGCAAGCAAAAA
TGTTAGGCGTTTCCATTAAGAAAATGCGTTGGTTTGTGTTTGTCTACTTTGTGCATTGCTTGTAGCAACACA
AGTCGCTGTTAGTGGGAGTATTGGGTGGATAGGGCTTGTATTCTCATTGACACGTTTTTTTGTAGGA
AGTGATCACCGTTATCTATTGCCCCCTCCTTTTTGATTGGTGGGATTTTCATGATTGTTATTGATACAC
TTGCACGTACGTTAACTTCTGCAGAAATTCCTGTAGGTATTATCACCGCTCTTTTAGGAGCACCCATTTT
20 TACCTTGCTCCTATTAAAACTTATCGAAAGAAGTCATTATGA

SEQ ID NO:12 polypeptide sequence of Orf6

MKSMLANQRGFITSLIFILFIIVLFTLNIGTFSLSSTGKVMSSILSKPFLSQHASFTPMEXHIVWHVRLPRIIM
AFFSGGIXAMSGATLQGVFHNPLVDPHIIGVTSGAVFGGSLAILLGFPYLLILSTFSFGLLTLFLIYVTM
FIGKGNRIVLVLAGVILSGFFSTLVSLIQYLADAEVLPISIVFWLLGSFATTSWAKLAILLPCVFIAAYLLF
25 RLRWHINVLISLQDMQAKMLGVSIIKKMRWFVLLLCALLVATQVAVSGSIGWIGLVIPHLTRFFVGS DHRVLLP
ASFLIGGIFMIVIDTLARTLTSAEIPVGIITALLGAPIFTLLLLLKTYRKSL.

SEQ ID NO:13 polynucleotide sequence of Orf7

ATGATTCAACGCTACGTTAAAATAGTCAGTATTGCTTTTATTACTTTTCTTAGGTTCTATTAATAATGCGT
TTGCAGCACGTGTTATTACTGATCAATTAGGACGAAAGGTCACATATCCAGATGAAGTTAATCGTGTGTTGT
30 TGTCTGACAGCATCAGACTTTAAATCTCCTTGCCCAGCTTGATGCAAAGGAAAGTGTAGTCGGAGTGTTA
TCAAGTTGGAAAAACAATTAGGGAAAACTATGCACCAAAAGAAATGATTGAGCAAATCGAACAGGCTG
GTGTGCCTGTTGTAGCCATTTCTTTGCGTGAAGATAAAAAAGGTGAAGAAGGAAAAGTCAACCCAGAAAT
GGAAGATGAAGAAGTTGCCTATAATAATGGTTTGAAACAAGGCATTTATTTAATTGGTGAAGTAATTAAT
CGACAAGCGCAAGCCCAAAAGCTAGTTACTTACACTTTTGAACAGCGTGAATTAGTGAGTCAACGTTTAA
35 GTAAGGTGCCTGATGAGCAGCGTGTAGGGTCTATATTGCAAATCCAGATTTAGCGACTTATGGTTCTGG
AAAATATACAGGGTTAATGATGCTTCATGCTGGAGCGAAGAATGTGGCAGCTGAAACAATAAAAGGTTTT
AAACAAGTTTCGATTGAGCAAGTGATTCAATTGGAATCCTGCAGTTATCTTCGTACAGGAACGTTATCCTC
AGGTTATCGAGCAAATTAATAAGGATCCCTCTTGGCAAATTATTGATGCGGTGAAAAATCAACGTATCTA
TTTAATGCCGGAATATGCAAAAGCGTGGGGATATCCAATGCCTGAAGCATTAGCGATTGGTGAATTATGG
40 TTAGCAAAACAACTTTACCCTGAATTGTTTGCAGATGTTGATTTAGAGGAAAAAGTAAACCAATACTATA
AATTGTTCTATCGTATGCCATATAACCAAGTAA

SEQ ID NO:14 polypeptide sequence of Orf7

MIQRYVKIVSIALLLFLGSINNAFAARVITDQLGRKVTIPDEVNRVVVXQHQTNLNLAQLDAKESVVGVLSS
WKKQLGKNYAPKEMIEQIEQAGVPVVAISLREDKKGEKGKVNPEMEDEEVAYNNGLKQGIYILIGEVINRQAO

B45292

AACGCCGGATTTTTAGCCCAAACCGAGCAAGAAATTACCGCTTGGTGCGAAGCGCAGGGCATAGCCTTAA
ACAACAAAAACAAGACCAAGCTGCTGGACGTCAAAACCTGGGAAAAAGCCGCCGCACTTTTTTCAGACGGC
ATCAACCTTGCTCGAACATTTTCGGCGAACAACAATTTGACGATTTCAACCAATTCAAACAAGCCGTGGAA
TGCCGCTCTGAAAGCCGAAAAAATCCCCCTTTCTGCCACAGAGAAAAAGGCCGTTTTCAATGCCGTAAGTT
5 GGTACGACGAAAATTTCAGCCAAAGTGATTGCCAAAACACTCAAGCTCAAACCAAACGAATTGGACGCCCT
TTGCCAACGCTACCAATGCCAAGCCGACGAGCTGGCAGACTTTGGCTATTACGCCACCGGCAAAGCAGGC
GAATATATCCTATATGAAACGAGCAGCGACTTGCGCGACAGCGAATCCATACCGCTCAAACAAAATATCC
ACGACTATTTCAAAGCCGAAGTGCAAGCGCACATCAGCGAAGCATGGCTGAATATGGAAAGCGTAAAAAT
CGGCTATGAAATCAGCTTCAACAAATACTTCTACCGCCACAAACCATTACGCAGCCTTGCGAAGTTGCCCA
10 AGATATTTTGGCGTTAGAAAAACAGGCTGACGGCTTGATTAGTGAAATTCTAGAGGCTTAA

SEQ ID NO:18 polypeptide sequence of Orf9

MEHSVHNLKVSFIWSIADDCLRDVYVRGKYRDVILPMFVLRRLDTLLEPSKDAVLEEMRFQKEELAFTELDD
LPLKKITGHVYNTSKWTLKSLYQTASNTPOYMLANFEEYLDGFSTNIHEIINCFKLREQIRHMSHKNVLLS
15 VLEKFVSPYINLTPKEQQDPEGKLPALTNLGMGYVFEELIRKFNEENNEEAGEHFTPREVIELMTHLVFDP
LKDQIPAIITIIDPACGSGGMLTESQNFIEQKYLSESGERSIFLFGKETNDETYAICKSDMMIKGDNPEN
IKVGSTLATDSFQGNHDFMLSNPPYGKSWSKDQAYIKDGNEVIDSRFKVTLDPDYWGNVETLDTATPRSSDGQ
LLFLMEMVSKMKS PNDNKIGSRVASVHNGSSSLFTGDAGSGESNIRRHIEKDLLEAIVQLPNNLFYNTGITT
YIWLLSNNKPEARKGKVLIDASLLFRKLKRLGDKNCEFVPEHIAEITQNYLDFTAKARETDSQNEAVGLA
20 SQIFDNQDFGYKVITIERPDRRSAQFTAENISPLRFDKALFEPMQYLYRQYGEQIYNAGFLAQTEQEITAWC
EAQGIALNNKNKTKLLDVKTWEKAAALFQTASTLLEHFGEQQFDDFNQFKQAVECRLKAEKIPLSATEKKAV
FNAVSWYDENSARKVIAKTLKLPNELDALCQRYQCQADELADFGYYATGKAGEYILYETSSDLRDSSEIPLK
QNIHDYFKAEVQAHISEAWLNMESVKIGYEISFNKYFYRHKPLRSLAEVAQDILALEKQADGLISEILEA.

SEQ ID NO:19 polynucleotide sequence of Orf10

ATGCAGCCGGAAAAACCAATATTTTGGACGCAAAGGACTAGGAGAAAAAGACATCAAGCCAACTAAAATAG
25 CTGAAGAATTAGTTGGAATGCTCAATGCTGATGGCGGAGTTTTGGCTTTTGGTGTGGCAGATAATGGCGA
AATCCAAGACTTGAATAGCCTTGCGGATAAATTAGATGATTATCGGAAATTGGTTTTTCGATTTTATTGCA
CCGCCTTGTGGATTGGACTGGAAGAAATCTGGTTGATGGAAAATTAGTTTTCTTATTCCACGTAGAGC
AAGATTTAGAGCGTATTTATTGTGCGCAAAGACAATGAAAATGTGTTCTTACGTGTAGCAGATAGTAATCG
AGGCCCTCTCACCAGAGAACAAATCAAAAATCTTGAATATGATAAAAATATCCGCTCTATTTGAAGATGAA
30 ATAGTTCCTGATTTTAATGAAGAAGATTTAGATCAAGAATTATTAGAGCTATATAAAAAAGAAAGTTAATT
TTACCTCCGATAATATCTTAGATTTATTATACAAGCGAAATTTATTAACCAAAAAGGAAGTTGTATCA
GTTTAAAAAATCAGCCATTTTACTCTTTTCTACCATGCCGGAACGTTACATTCCTTCAGCATCAGTCCGC
TATGTTTCGTTATGAAGGTACAGTAGCGAAAGTCGGTACTGAGCATAATGTGATAAAAAGACCAACGTTTTG
AAAATAATATTCCAAAGCTAATTGAGGAGCTGACCTATTTTTTAAGAGCCTCTTTAAGGGATTATTACTT
35 TCTTGATGTCAATCAGGGAAAATTTATCAAAGTACCGGAATATCCTGA

SEQ ID NO:20 polypeptide sequence of Orf10

MQPENQYFERKGLGEKDIKPTKIAEELVGMLNADGGVLAFGVADNGEIQDLNSLGDKLDYRKLVFDFIAPP
40 CRIGLEEILVDGKLVFLFHVEQDLERIYCRKDNENVFLRVADSNRGPLTREQIKNLEYDKNIRLFEDEIVPD
FNEEDLDQELLELYKKKVNFTSDNILDLLYKRNLLTKKEGQYQFKKSAILLFSTMPERYIPSASVRYVRYEG
TVAKVGTEHNVIKDQRFENNIPKLEELTYFLRASLRDYYFLDVNQKFIKVPPEY

SEQ ID NO:21 polynucleotide sequence of Orf11

ATGTCAATCAGGGAAAATTTATCAAAGTACCCGGAATATCCTGAAGAAGCTTGGTTAGAAGGTGTTGTAA
ATGCGCTTTGTGCATCGTTCTTACAATGTTCAAGGTAATGTTATTTATATTAAACATTTTCGACGATCGTCT
45 TGAAATTAGTAATAGTGGCCCTCTCCCTGCTCAAGTCACCATTGAAAATATTAAAACGGAACGATTTCGCT

B45292

TTTTGATTTGCTCTACCCCGTTCCGCTTGCCAGCAGCGGCGAAAAGGTCAAGCAGCGTTTTGAACAGAAT
TTGTTTAGCTGTATGCGTCAAGTGCTTATTCTGCCTCAAGCAATGAAACGGTGGATATGGTGCTGTTTG
CCAATGGCTTGCCGATTATTGCCCTTGAGCTGAAAAACCATTGGACAGGTGAGACAGCCATTGATGCGCA
5 AAAACAATACCTCAACCGTGATTTAAGCCAAACGTTGTTCCATTTGCGGCGTTGTTTGGCGCATTTTGCC
TTAGATACGGAAGAAGCTTATATGACCACCAAATGGCGGGGCTGCTACGTTTTTCTTGCCGTTTAACT
TGGGCAACAACCTGCGGTAAAGGTAATCCGCCCAATCCCAATGGACACCGCACGGCGTATTTATGGCAAGA
GGTGTTGCGCAAAGCAAGCCTTGCCAACATTATTAGCATTTTATGCGCTTAGACGGTTCAACCAAAGAT
CCGTTGGATAAACGTACCCTCTTTTCCCTCGCTATACCAATTAGATGTGGTCCGCCGTTTGATTGCTG
10 ATGTCAGTGAACATGGCGTGGGTAAACGTTATTTGATTCAACATTCTGCCGTTTGGGCAAGTCTAATTC
CATTACTTGGCTGGCGTATCAGTTGATTGAGGCATATCCGCGCAATGAAAAGGCGGCAAACGGTAGAGAG
GCAGACCGCCCGATTTTGTGATTGCGTGATTGTCGTAACCGACCGTCGTTTGTGGATAAGCAACTGCGCG
ACAATATCAAAGATTTTTCAGAAGTTAAAAACATTGTTGCGCCGGCGTTGAGTTCCGGCAGAGTTGCGCCA
ATCGCTTGAGCAGGGCAAAAAAATCATTATTACCACGATTCAAAAATTCCCGTTTATTGTGCGATGGCATT
15 GCTGATTTAGGCGACAAACAATTTGCGGTGATTATTGATGAGGCACACAGCTCACAATCAGGTTCCGGCAC
ACGACAATATGAACCGGGCCATCGGCAAAACGGAAGACCTTGATGCTGAAGATGTGCAAGATTTGATTTT
ACAAACCATGCAATCCCGCAAAATGCACGGCAATGCGTCGTATTTTGCTTTCACCGCCACACCGAAAAAC
AGCACTTTGGA AAAAATTCGGCGAAAAACAGGCGGATGGCAAGTTTAAAGCCGTTCCACCTTTATTCTATGA
AGCAGGCGATTGAAGAAGGCTTTATTTTGGATGTAATCGCCAATTACACCACCTATAAAAGTTTTTATGA
GATCACTAAGTCGATTGAAGATAATCCGGAGTTTGATAGTAAAAAGGCTCAAAGCCGTCTGAAAGCCTAT
20 GTGGAGCGTTCGCAACAAACGATTGATACTAAAGCGGAGATAATGCTGGATCATTTTATTTACCAAGTTT
TCAACCGTAAAAAACTCAAAGGCAAAGCCAAGGGAATGGTGGTAACGCAAAATATTGAAACCGCCATCCG
CTATTTTCAGGCGTTAAACATTTGCTGGCCGGGCGGGGTAATCCGTTTAAATTTGCGATTGCGTTTTCA
GGCAGTAAAGTGGTTGACGGTGTGGAATACACCGAAGCGGAAATGAACGGCTTTGCAGAAAGCGAAACCA
AAGAGTATTTGATCAAGATGAATATCGTTTGCTGGTGGTGCAGCAATAAATATCTGACCGGTTTCGATCA
25 GCCGAAATTGTGTGCCATGTATGTGGATAAGAACTCTCCGGCGTGCTTTGCGTGCAGGCTTTATCTCGT
TTGAATCGCAGTGCGAATAAGTTGAGTAAACGCACGGAAGATTTGTTTGTATTGGACTTTTTTAACAGCG
TTGAAGATATTCAGCAGGCATTTGAGCCGTTTTATACTTCTACTTCGTTGTGCGAGGCAACCGATGTCAA
TGCTTTCATGATTTGAAAGACCGGTTGGATGAAACCGGCGGTGTACGAACAAGCGGAGGTCAACGATTTT
ACTGAAGGCTATTTTGCCAATAAAGACGCACAGCAATTAAGCAGTATGATTGATGTGGCTGTCCAACGTT
30 TTGATGATGAATTGGAATTGGATTTGGATCGAAATGAAAAAGTTGATTTTAAATCAAGGCAAAACAGTT
TTTAAAAATTTACGGGCAAATGGCCTCCATCATCAATTTTGAAAATATCGCTTGGGAAAAGCTCTATTGG
TTCCTCAAATTCCTTAGTACCCAAATTAAAAGTACAAGACCCGATGGATGAATTTGATGAAATTTTAGATG
CAGTGGATTTAAGCTCTTACGGCTTGCGGCACACCAAGCTGAATTACAGCATTAAATTAGATGATGAAGA
AACAGAGCTTGACCCGCAAAACCCCAATCCGCGCGGTACGCATGGTGAAGATAAAGAAAAAGATCCGATT
35 GATGAAATTATTCTGTGATTTAACGAAAGATGGTTTCAAGATTGGAGCGCAACGCCGGATGAGCAACGGG
TAAATTTTATCAATATTACCGAGCGCATCCGCGAGCCATAAAGACTTTGAGCAGAAATATCAAAATAACCC
GGATATTCATACCCGTGAATTGGCTTTCCAAGCCATTTTGCGCGATGTGATGAGCGAACGCCATAGGGAT
GAATTAGAGCTATACAACTTTTGGCCAAAGATGCCGCATTTAGAACCGCTTGGACGCAAAGTTTGCAAC
GGGCTTTGGCTGGATAG

SEQ ID NO:26 polypeptide sequence of Orf13

MVSGTKEKDLEIAIEKALTGTWRENMENKLGEPKAEYLPRHHGFKLAFSQDFDAQFAIDTRLFWQFLQTSQE
AELARFQQLNPNWDQRKILERLDRQIKNGVLHLLKKGLDIDSAHFDLLYPVPLASSGEKVKQRFQNLFS
MRQVPYSASSNETVDMVLFANGLPIIALELKNHWTGQTIDAQKQYLNRLDSQTLFHFGRCLAHFALDTEEA

B45292

MSEYKLNPPPTVSSY TENMMLKVLF EHKGFSEV FRETSWRSDEIASAFGLPEELEN DKNLRTVARRLKERYK
KLOKSTALLPELWKQAYENLATLAEFLQ LNPVEQELLRFAMHLRSEGAMRDLFGYLPKSDLQRTAAIMADLL
KQPKNQILSALKKGSKL DAYGLIDRDYRPSVHDYLDWGETLDFDEFVTQPLNENVLLKSCTEVAQVPSLQL
DDFDHIAGMKEMMLTYLQQALKHHRKGVNLLIYGVPGTGKTEFAGLLAQALGISAYNITYMDSGDVVEAEQ
5 RLNYSRLAQ TLLNGKQALLIFDEIEDVFNGSFMEERSVAQKNKAWTNQ LLENNVPMIWLNSVSGIDPAFLR
RFDFILEMPDLPLKNKSALITQLTEGKLSPAYVQHFAKVRSLTPAILSR TIRVAKELNTSNFAETLLMMFNQ
TLKSQNKPKIEPLVLGKADYNLDYVACNDNIHRISEGLKRSKKGRICCYGPPGTGKTAWAAWLAEQLDMPLL
LRQGSDLLNPNYVGGTEQNIQAFAEQAKADNAILVLDEVDTFLFSREGANRSWERSQVNEMLTQIERFEGLMV
10 VSTNLIEVL DHAALRRFDLKLKFDYLT LKQRLDFAKQQA EILGLPLLSEEDLSQIESLNLITPGDFAAVARR
HQFSFPFKVQDWLMALQGECEVKPAFSATTRRIGF.

SEQ ID NO:29 polynucleotide sequence of Orf15

ATGTTTGAAAAAATTGAACCTACTAATATTCGTTTTATTAAATTAGGCATAAAAGGATGTTGGGAAAAAG
ATTGTATTGATAAAAATAGTACAGCAAGTACAAAAAATACGATTCGTCTTGCTATGAATCTACATCAGA
GATTCACAAAGAATGTTTGAATAATCAATGGGATAGTTGTATTGAATATTGTA AAACCTATTGGAGTGAC
15 CATACAGGAAC TGTTC AAATCACTTGAGACAAATTC AAGATTTTTATCAACTTGGGGAAGATACACTTT
GGATCACCTTCTTTGGACGTAAATTATATTGGGCTTTTTTG CAGTAAAGAGGTTGTTGAGGAAAGCGATGG
TTCTAGAACAAGAAAAGTTATTAGTAACAATGGGAATTGGTCTTGCGTTGATGCTAACGGTAAAGAGCTT
TTAGTCGATAATCTTGATGGTAGAGTAACAAAGGTCCAAGCCTATAGAGGGACGATTTGTGGTGTGAGA
TGGAGGACTATTTAATACGTCGTATAAATGGTGAAGTTATTGAGGAAATTACAGAAGCGAAAGAGGCGTA
20 TGAACATTAATTAAATCAGTTGAAAAATTAATTAAAGGTTTATGGTGGAGTGACTTTGAAC TTTTAACG
GATCTTGTTTTTTCTAAATTAGGATGGCAACGATACTCTGTTTTAGGTAAAACGGAGAAAGGAATAGATC
TTGATTTGTATTTCGTCTTCAACGCAGAAAGAGAGTATTTGTGCAAATTAAGTCAGATACGGATATTAAACA
ATTAGACGAATATGTTTTCGAAC TTTGAAAGTGAATATAAAAACTATGGTTATT CAGAAATGTATTACGTA
TATCATTCTGGTTTAGAAAACATAGATGAAAAACAATATCAAGCTAAAGGAATTAAGCTTGTAATGGCC
25 GAAAAATGGCAGAGCTTGTAATTAGTGTGCTGGTTTAGTTGAATGGTTGATTAAACAAACGTTCTTAA

SEQ ID NO:30 polypeptide sequence of Orf15

MFEKIEPTNIRFIKLGKGCWEKDCIDKNSTASTKNTIRLGYESTSEIHKECLNNQWDS CIEYCKTYWSDHT
GTVSNHLRQIQDFYQLGEDTLWITFFGRKLYWAFCSKEVVEESDGSRTKVISNNGNWS CVDANGKELLVDN
30 LDGRVTKVQAYRGITCGVEMEDYLIRRINGEVIEEITEAKEAYETLIKSV EKLKGLWWSDFELLTDLVFSK
LGWQRYSVLGKTEKGIDLDLYSSSTQKRQVQIKSDTDIKQLDEYVSNFSEYKNGYGYSEMYVYHSGLENI
DEKQYQAKGIKLVNGRKM AELVISAGLVEWLINKRS.

SEQ ID NO:31 polynucleotide sequence of Orf16

TTACCCCTTTGCCAACAAAATTGGCAGCAACAAGCGACGCAACCAAGATGCCCTTTTTTAATGGCGAGGCGG
TGTTTCAATATAAACTCAAAACGGCTGAAAAACGCCTTGAAAAACCGACCGCACTTTATTGTGGGCGTGCG
35 AGATGGTATTTCTAATAGCAACCGACCTGAAAAAGCGAGCAAATTTGGCTATGCAATTATTAAGCCAAATG
GAAAGTATAAACCGTCAAACGATCTACGATTTACAATCCAGTTTATCAGCAGAATTAGCTGAGGATTATTT
TTGGTTCTGGCGACCACATTTGTGGCTGCCGAAATTGATCAAATAACCCGTAAAGCGAAAATTCTCAGCGT
AGGCGATAGTCGTGCTTATTTAATTGATGCCCAAGGAAAATGGCAACAAATCACCCAAGATCATTTCTATT
CTTTCTGAATTATTGACTGATTTCCCCGATAAAAAAGAGAAGATTTTGCCACGATTTATGGCGGCGTTT
40 CTTCTTGTTTTAGTCGCCGATTATTCGAATTTCAAGATAAAATTTTTTATCAAGAAATTGAAATTCAGCA
AGGGGAAAGTTTATTACTTTGTTCTGACGGCTTGACCGACGGGCTTTCAGATGAAATGCGCGAAAAAATT
TGGCAGAAATATCCCGATGATAAATATCGCCTTACGGTTTGCCGCAAGATGATTGAGAAGCAATCGTTTT
CGGATGATTTGTGGTAGTTTGTGTGTCATTCTATTATTAGTAA

SEQ ID NO:32 polypeptide sequence of Orf16

LPFANKIGSNKRRNQDALFNGEAVFQYKLKTAEKRLNRPHFIVGVADGISNSNRPEKASKLAMQLLSQMES
45 INRQTIYDLQSSLSAELAEDYFGSATTFVAAEIDQITRKAKILSVGDSRAYLIDAQGWQQITQDHSILSEL

B45292

TCAATGATACTGGGCTTGCCGCGCTCACTTTATTAGTTGCTGAATCTGATCCGAAACAAAAGAAACGCT
TATTAGGCTTATTATGCATATGCTTAAGCAAGAGAAAAATGA

SEQ ID NO:38 polypeptide sequence of Orf19

MLVIKENNMNNQNPIEIIYQTDGTTQVEVRFENDTVWLSQAQMAMLFKDIRTINEHITNIFDDEELEKE
5 STIRKFRIVRQEGKRQVNREIEHYDLDMIISVGYRVKSKQGISFRRWATARLKEYLTQGYTINQKRLQQN
AHELEQALALIQKTANSSELTLESGRGLVDIVSRYTHTFLWLQQYDEGLLAEPQTQGGTLPITYAEAFSA
LAELKSQMLTKGEASDLFGRERDNLSSAILGNLDQSVFGEPAYPSIEAKAAHLLYFVVKNHFFSDGNKRS
GAFLFVDFLHRNGLFDHNGYPVINDTGLAALTLVAESDPKQKETLIRLIMHMLKQEKK.

SEQ ID NO:39 polynucleotide sequence of Orf20

10 ATGACAGAGAAAAATAAACCAATTTGCGTGGTATTAACGGGAGCTGGCATTAGTGCCGAAAGTGGAATTC
CAACTTTTAGATCGGAAGATGGTTTGTGGGCAGGGCATAAAGTAGAAGAAGTTTGTACGCCCGAAGCCTT
GCAAAAGAACCGTGCGAAAGTGCTTGATTTCTATAACCAACGCCGTAAAAATGCGGCAGCAGCTAAGCCA
AACGCTGCGCATCTCGCCTTAGTTGAAC TAGAAAAAGCCTATGATGTGAGAATCATCACGCAAAATGTGG
ATGATTTACATGAACGTGCCGGCAGCTCGAAGGTGTTGCATTTACACGGTGAATTAAATAAAGCTCGCAG
15 TAGCTTTGATGAAAGTTATATTGTGGATTGTTTTGGTGATCAGAAATTAGAAGATAAAGATCCAAATGGA
CACCCAATGCGCCCTTACATCGTCTTTTTTGGTGAAATGGTGCCGATGCTAGAACGAGCGGTTGATATTG
TGGAACAAGCAGATGTTGTGTTAGTGATTGGCACTTCTTTACAAGTGATCCAGCCAATGGCTTAGTCAA
TGAAGCCCCAAGAAAAGCGCCAATTTATCTGATTGATCCTAACCCAAATACAGGATTTGTTCTGTAAGCAA
GTTATTGCAATCAAAGAAAAGCAGGCGAGGGTGTGCCAAAAGTGGTGGCAGAGTTATTAGAGAACACCA
20 AAAACTCATAG

SEQ ID NO:40 polypeptide sequence of Orf20

MTEKNKPICVLTGAGISAESGIPTFRSEDGLWAGHKVEEVCTPEALQKNRAKVLDIFYNQRRKNAAAAKP
NAAHLALVELEKAYDVRIITQNVDDLHERAGSSKVLHLHGELNKARSSFDESIVDCFGDQKLEDKDPNG
HPMRPYIVFFGEMVPLERAVDIVEQADVVLVIGTSLQVYPANGLVNEAPRKAPIYLIDPNPNTGFVRKQ
25 VIAIKEKAGEGVPKVVAELLENTKNS.

SEQ ID NO:41 polynucleotide sequence of Orf21

ATGAAGAAAATTGTTTATATTGATATGGATAATGTGATGGTAGATTTTCCATCAGGTATTGCAAAACTAG
ATGATAAAACCAAGCGAGAATATGAAGGTCGATATGATGAAGTCGAGGGCATTTTTAGCTTAATGGAACC
TATGCCGAATGCGATTTCTGCGGTGCATAAATTGATGAAAAATATCATATTTATGTGCTTTCTACTGCG
30 CCTTGGCATAATCCTTTTGCTTGGAGTATAAAAGTAAAATGGATTACCATTAATTTTCGGTGAAAGAAAAG
GTTTCAGCCTTATATAAACGATTGATTTTATCCCATCATAAAATCTCAACCAAGGTGATTATTTAATTGA
TGATCGCACTAAAAATGGTGTGGCAAATTTCAAGGCGAGCATGTTTATTTTGGTACAGAACAGTTTGCT
AATAAAAGGAGCCTGAAAAATGACAGAGAAAAATAA

SEQ ID NO:42 polypeptide sequence of Orf21

35 MKKIVYIDMDNMVDFPSGIAKLDDKTKREYEGRYDEVEGIFSLMEPMENAI SAVHKLMKKYHIYVLSTA
PWHNPFAWSIKVKWIHHYFGEEKGSALYKRLILSHHKNLNQGDYLIIDRRTKNGAGKFQGEHVHFGTEQFA
NKRS LKNDREK.

SEQ ID NO:43 polynucleotide sequence of Orf22

40 CATTATCGGAGTATTCACGGTAAAGAACATAAGGCACAGGTCAAGCCCTTGGCTTTGGTTCAACAAGGAC
CAAGTAGCTATTTAGTCGCACAATATGAGAATGGCGATATTTTACACCTTGCTTTGCATCGCTTGCTTAA

B45292

AATGAGCGTTTCGAGCTCCTGTTGGGGACATTAATATAGCACTTGAAAAATGCTGTATTGGTCGCGGATTA
GCTGCATTACAACATAAGAGTAAAAGTTTGTCTGTTTCGGTTTATATCAAATACAATCTATAAAACCAGAAT
TAGATTTATTTAATGGTGAAGGAACTGTTTTTGGTTCTATCAATCAGGATAACTTAAAAAATATCCAAAT
TATTAACCTGATGAAAAATTTATTCAGCTTTTTGAAAAATATTTATCATCTTGTGATTCAAAAATTATG
5 AATAACGAGATAGAAAAATAATGCACCTGAAAGAAATAAGGGATTTATTGTTACCTAGATTATTGAGTGGAG
AAATTCAATTATGA

SEQ ID NO:48 polypeptide sequence of Orf24

MNDWKVITLADCASFOEGYVNP SKNEPSYFGGTIKWLRATDLNNGFVYKTSQTLTEKGFLSAKKSAVLFEPD
10 SLAISKSGTIGRIGILKDYMCNRAVINIKVNENICNPLFI FYTLLNSKEQIETLAEGSVQKNLYVSALSKV
KLLLLLDINKQKEIGYIILNTLDQKIELNTQINQTL EQIAQALFKSWFVDFDPVRAKIQALSDGLSLEQAELAA
MQAISGKTPEELTALSQTQPDYAEAEAKAFPCMEVEVDGVEVPKGWELSTIGDCYDVVMGQSPKGETYN
ENKQGM LFYQGRAEFGWRFP TPRLF TTDPKRIAEQNSILMSVRAPVGDINIALEKCCIGRGLAALQHKSKSL
SFGLYQIQSIKPELDL FNGEGTVFGSINQDNLKNIQI INPDEKFIQLFEKYLSSCD SKIMNNEIENNALKEI
RDLLLLPRLLSGEIQ L.

SEQ ID NO:49 polynucleotide sequence of Orf25

ATGGAATTAATAAGCGATAATCCAATAAAAGATTCTAGCAATGATTTATTAGGTAGAGCTAGTAGTGCAG
AAGCATTTTGCTAAACACATTTTTTTCATTTGACTATAAAGAAGGTTTGGTTGTGGGATTATGTGGAGAATG
GGGAAATGGTAAACATCCTATATAAATTTAATGCGACCAGAATTAGAAAAAATTCTTTTGTACTTGAT
TTTAATCCTTGGATGTTTAGTGATGCTCATAACTTAGTTGCTTTATTTTTTACTGAAATCTCTGCTCAGT
20 TAAGAGATTATGAGGATGATAATGAGCTAATTGATAGTTTGGAGAGTTGTTATCTAATTT
AAAACCTATTCCATTTGTAGGAAATTATTTTAGTGCTTGGGTGGCTGTTTAAGTTTTTTTTCAAAGAAA
AAGAAAGAAAAAACAGTTTGAAAAATCAACGTGATAAATTAATTAAAGTTCTAAAGGAAATAAGTAAAC
CTATTACTGTAATTTTAGATGATATAGACCGTTTATCATCTGATGAATTACAATCAATTCTAAAATTGGT
CAGAGTTACAGGAACTTTCTTAATATTGTTTATGTTTATCATTTGATAAAAATAGAGTAATTAAACCA
25 TTAAATGATAATACCATTGATGGCCAGGATTATTTAGAGAAGATAATTCAGATTCCATTCGATATACCAC
AGGTACCTAAAAAATATTACAAGAAAATTTATTTTTCATCTTTAGATAAGATTTTAAGGGATGTTTACCT
AGATAAGGCGCGTTGGTCTAATGCATATTGGAATATCATTAAGCCAACAATAAAAAATATTCGAGATATT
AAGCGTTACACATCTTCTCTATCGAATATCTTTAAACAATTAGGTAAAGAAAATTGATGTGGTTGATTTAC
TCACTATTGAAGCGATAAGAATTTTCTTTCCAGATAAATTTAAAGAAAATTTTGAACCTAAAGATTATCT
30 CTTGGCAGCATCAGATAATGACAAAAGAAAAGTTAAGTTAAGTGATTTTATTCAAGATAATGAAATGTAT
GAGTCTTTTCTAGAAGTTTTTATTGATATTGATAATATAAATTCAAATAATGAATTCCTAAAAAATAGAA
GGATTGCTTATTCGGCATTCTTTGATTTATATTTTGAACAAGTTATGAGTCCTGAGTTCATAAATGTTAA
ATTATCACAAAAGTTTGGCTTGCAATGCAGTCAGAAGAAGATTTCAAGATCGCTTTATCAGCTGTTTCTT
GACGATTCTCTAGAAAATGTAGTTAACAATTTAATTGACTATGAAAAAGACTTTACTAAAGAAATAGCTC
35 TAGCAACTATACCAACATTATATAGAAATTTACCAAGAGTGCCTGAAAAAGAATTAGGATTCTTTGACTT
TGGGGCGGATATGGTTTGGAGTCGCTTAGTTTATAGATTACTTAGAAGACTTCCTGAGAAGGATAAAAAA
GAAGTTATTACTCAACTATTAAATTCTAGCGATCTATATGGGCAATATCAAATTGTAGGAATTATTGGAT
ATCGAGAGGGCCGAGGTCAATTAGTATCTGAATCGGATGCAAAGACTTGGAGGAAATATTTTTTAA
TAATATTCGCTCTGCAACAATTAAAGAAGCTTGCAAGGAACCTATAATTTGTACATATAATCTATTTCTTT
40 GTTTCAATTGGAAACCTTTTTCTGATGATATATTAAGTTCCCTGAAGTATTTTTATCATTACTTAAAT
CTTCAATATCAGAACGTAAATCTCAAAGAGGGGATGATCCTACAATACATAGAGAGAAAATTCTACTTTG
GGATGCCTTAATTAATAATTTGTGGAGATGAGGATAAAGTAAATAGTTTAATTGAAAAAATAGCTGAAGAT
GAAGAACTTAGAAATAAAGATTATATGGAACCTGCAATTAATATAAGAATGGATACCGACATAAAAAAT
CAATGAATCATGAAGATGATTTAGATGAGTTTAA

B45292

CGCGAAATTCGAGCCAGTGAATTGCGGTTTTATCAAAAAGTACGAGAGTTATTTAAATTATCCAGTGACT
ACGATAAAACAGATAAAAGTCACTCAAATGTTTTTGCAGAAACACAAAATAAGTTGATTTATGCCATTAC
ACAACAAACCGCCGAGAGCTTATTTGTACGCGTGCAAATGCCAAATTGCCTAATATGGGTCTTACCTCT
TGGAAAGGTGCTGTTGTACGTAAAGGCGATATTATTACCGCTAAAAACTATTTAACTCATGATGAATTAG
5 ATTCTTTGAATCGTTTAGTGATGATCTTTTTAGAAAGTGCTGAATTACGCGTTAAAAATCGTCAAGATCT
CACATTAAATTTCTGGCGTAATAATGTCGATAATTTAATTGAATTTAACGGTTTTCCGTTGCTTATCGGT
AATGGAACCCGAACCGTAAAACAAATGGAAACCTTTACCAAAGAACAATATGCCTTATTTGATCAGGTCA
GAAAAACAACAAAAACGCATACAAGCTGATAATGAAGATTTAGAAATTTAGAAAACCTGGCAGAAAGATCT
GAAAAAGCAAAAGCATTAA

SEQ ID NO:54 polypeptide sequence of Orf27

MNDLIYNTDDGKSHVALLVIENEAWLTQNQLAELFDTSVPNITTHIKNILQDKELDEFVSIKDYLIITAQ
DSKQYQVKHYSIDMILAIGFRVRSRPGVQFRRWANTQLRTYLDKGFLDKERLKNPQGRFDHFDELLEQI
REIRASELRFYQKVRELFKLSSDYDKTDKVTQMFFAETQNKLIYAITQQTAAELICTRANAKLPNMGLTS
WKGAVVRKGDIIITAKNYLTHDELDSLNLRLVMI FLES AELRVKNRQDLTLNFWRNNDNLIEFNGFPLLIG
15 NGTRTVKQMETFTKEQYALFDQVRKQKRIQADNEDLEILENWQKDLKKQKH

SEQ ID NO:55 polynucleotide sequence of Orf28

ATGCAACAGCGTGCTACTTTTTTTAAAAGCGTGCTAAGCCAACGTTATACTAAAACCTGAACTGTGTCAGC
AGTTTAAATATTAGCCGTCCAACGGCAGATAAATGGATTAAACGCCACGAACAGCTTGGTTTTGAGGGCTT
AAGCGAGTTATCTCGTAAATCTTATCATAGCCCTAATGCCACGCCACAATGGATTGTGACTGGCTTATC
20 AGTGAGAAACTTAAACGTCCTCACTGGGGTGCCAAAAAGCTTTTAGATAACTTTACTCGGCATTTTCCAG
AAGCGAAAAAGCCGTCTGATAGCACGGGCGATTTAATTTTGGCGTGTGCAGGGTTAAAACGTCGTATGAG
TGCAGACACACAATCTTTTGGCGAATGCATCGCACCCAATACCACCTGGAGTGTGCTGACTTCAAGGGGCAA
TTTTTACTCGGCAATCAGAAGTTCTGCTATCCGCTGACGATTACAGATAATTTCAAGTCCGTTTTTATTTT
GTTGTAAGGGGTTGCCGAATACAAAATCAGCGCCTGTTATTGCTGAGTTTGAACGTCTTTTGGAGCAATT
25 TGGTCTGCCGTATTTCGATTTCGTACCGATAACGATTTCATCTTTTGCATCACAAGCATTAGGTGGATCTAGG
TGTATTGACTTAGGTATTCTTCTGAACGAATTAAGCCATCACACCCAGAGCAGAACCGACGACACGAGC
GAATGCACCGTAGCTTAAAAACAGCGCTTCAACCTCAAATAGCTTTGAAGCTCAACAGACATTCTTCAA
CCAATTCTTACGAGAATACAAAGAAGAAATGTTACACGAAGGCGTTTGA

SEQ ID NO:56 polypeptide sequence of Orf28

MQQRVLFLKAWLSQRYTKTELCQQFNISRPTADKWIKRHEQLGFEGLSLSRKS YHSPNATPQWICDWLI
SEKLKRPHWGAKKLLDNFTRHFPEAKKPSDSTGDLILACAGLKRMSADTQSFGECIAPNTTWSADFKGQ
FLLGNQKFCYPLTITDNFSRFLFCCKGLPNTKSAPVIAEFERLFEQFGLPYSIRTDNDSSFASQALGGSR
CIDLGIPSERIKPSHPEQNGRHERMHRSLKTALQPQNSFEAQQTFFNQFLREYKEEC SHEGV .

SEQ ID NO:57 polynucleotide sequence of Orf29

TGCCAAACGGCGAACAAATCCGCAGAATTAAGCAGCGTTGTGGCTATTCTCGCTTCATGTTTAAATCGGGT
TAACTTGGCAGAATGAACAATATAAGCAAGATAATGGCGTCAAGTTTACGTTATACGAAAATCGCCAAATT
GCACCACAAAGTCACCAATACCCACAAAAAAACTACTTGCATCAAATCCCACACCGAATCAGCAAAAAC
CACGCAATGATTTATATTGAGAGTTTGCAAGCAACAAATTACCAAGGAGATGCGGAAAATACAGTAAAAC
GCGAAACAAAAATCAGACTTAAACCGTTCAACTTCAGCACAAATCTTGGCATGA

SEQ ID NO:58 polypeptide sequence of Orf29

CQTANKSAELSSVVAILASCLIGLTWQNEQYKQDNGVKFSYTKIAKLHHKVTNTHKKNYLHQIPHRISK
HAMIIYESLQATNYQGAENTVKRETKIRLKP FNFSTILA

SEQ ID NO:59 polynucleotide sequence of Orf30

B45292

TACTTTCCCCTGCTGGTGTAGCACAATTACAACACATTGATATTAGTGAAGTGCCTGCCGCATTCAAAAA
AATCAATAACCAATTTATTGGGATTTTAATTGGTGTGATTTTTCAGCTGAACCTTTACAACCGTTTCTATCAA
GTTGAATTACCAAAGGCACCTTCGTTCTTTAGCGGAAAACGCCCTCGTCCCAATTTTGGTTTCTTTCGTGA
TGATCGCCGTATCATTTGCCTTACTCTATATTTGGCCTCATATTTTAAACGCTCTCGTTTCATTTGGTGA
5 ATCCATCAAAGATTTAGGTGCAGTAGGTGCGGGGATCTACGGTTTCTTCAACCGCTTATTAATTCCTGTA
GGCTTACACCATGCCTTAAACTCTGTATTCTGGTTTGTATGTAGCGGGTATCAACGATATTCCAAACTTCT
TGGGCGGCGCTAAATCCATTGCCGAAGGCACCTGCAACCGTGGGGCTAACTGGTATGTATCAAGCTGGTTT
CTTCCCTGTGATGATGTTTGGTTTACCAGGTGCTGCTCTTGCAATTTATCACTGCGCAAAACCAAACCAA
AAAGTACAAGTGGCCTCAATTATGCTTGCGGGTGCCTTAGCCTCTTTCTTTACAGGGATCACTGAACCGC
10 TTGAATTCTCATTTATGTTTCGTTGCACCTGTACTTTATGTATTGCATGCATTATTAACAGGTATCTCTGT
ATTCATTGCAGCTACAATGCACTGGATTGCAGGATTCGGATTTAGTGCAGGTTTAGTGGATATGGTACTT
TCTAGCCGTAACCCACTTGCCGTTAGCTGGTATATGTTACTTGTACAAGGTATTGTATTCTTTGCTATCT
ATTATTTTGTGTTCCGTTTGTGCAATTAATGCCTTTAATCTCAAAACGCTAGGACGTGAAGATAAAGCGGA
AACAGCTGCAGCCCCAACTCAAAGCGACCAATCTCGCGAAGAAAGAGCGGTGAAATTTATTGCTGCTTTA
15 GGTGGTTTCAGAAAACCTTCAAACCTGTGGATGCTTGTATCACTCGTTTACGCTTAACTTTAGTTGATCATC
ACAATATTAACGAAGATCAACTTAAAGCGCTTGGTTCAAAGGTAATGTAAATTAGGCAATGATGGATT
ACAAGTCATTTTAGGGCCTGAAGCTGAAGTTGTGGCAGATGCGATTAAAGCAGAATTAAATAA

SEQ ID NO:66 polypeptide sequence of Orf33

MSVLSYAQKIGQALMVPVAALPAAALLMIGYWIDPDGWGANSQLAALLIKSGAAIIDNMGLLFAVGVA
20 GLAKDKHGSAAALSGLVGFYVVTLLSPAGVAQLQHIDISEVPAAFKINNQFIGILIGVISAELYNRFYQ
VELPKALSFSGKRLVPILVSFVMIASVALLYIWPFIHFNALVSFGESIKDLGAVGAGIYGFFNRLLIPV
GLHHALNSVFWFDVAGINDIPNFLGGAKSIAEGTATVGLTGMVQAGFFPVMFGLPGAALAIYHCAKPNQ
KVQVASIMLAGALASFFTGITEPLEFSFMFVAPVLYVLHALLTGISVFIAATMHWIAGFGFSAGLVDML
SSRNPLAVSWYMLLVQGIVFFAIYYFVFRFAINAFNLKTLGREDKAETAAPTQSDQSREERAVKFIAAL
25 GGSSENFKTVDACITRLRLTLVDHNNINEDQLKALGSKGNVKLGNDGLQVILGPEELVADAIIKAEIK

SEQ ID NO:67 polynucleotide sequence of Orf34

ATGAAAACAACTTCTGAAGAATTAACGGTATTTGTGCAAGTAGTCGAAAATGGCAGTTTTCAGCCGTGCAG
CCAAGCAGCTATCAATGGCAAATTTCTGCGGTAAGTCGTGTGGTGAAAAGGCTAGAAGAAAAATTGGGTGT
GAACCTAATCAACCGCACTACTAGACAGCTTAGACTAACAGAAGAAGGCTTACAATATTTTCGTGCGGTA
30 CAGAAAATTCTGCAAGATATGGCTGCAGCTGAAGCTGAAATGTTGGCAGTGCACGAAGTCCCAAGGCA
TACTACGCGTAGATTTCAGCCATGCCGATGGTGTACATCTGCTAGTGCCACTGGCAGCAAAATTCACGA
ACGCTATCCGCATATCCAACCTTCGTTAGTTTCTTCTGAAGGCTATATCAATCTGATAGAACGCAAAGTC
GATATTGCCTTACGAGCTGGAGAATTGGATGATTCTGGGCTGCGTGCTCGTCATCTATTTGATAGCCACT
TCCGCGTAATCGCCAGTCCAGACTACTTGGCAAAACACGGCACGCCACAATCAACTGAAGCTCTTGCCAA
35 CCATCAATGTTTAGGCTTCACTGAGCCCAGTTCACTAAATACATGGGAAGTTTTAGATGCTCAAGGAAAT
CCCTATAAAATCTCACCGTACTTTACCGCCAGCAGCGGTGAAATTTTACGGTCATTGTGTCTTTCAGGCT
GTGGTATTGCTTGCTTATCAGATTTTTTGGTAGACAATGACATCGCTGAAGGAAAAATTAATCCCTTACT
TACTGAACAAACCGCCAATAAAACGCTCCCCTTCAATGCTGTTTACTACAGCGATAAAGCAGTCAACCTT
CGCCTACGTGTGTTTTTAGACTTTTTAGTAGAAGAGCTAAGGGGATAA

SEQ ID NO:68 polypeptide sequence of Orf34

MKTTSEELTVFVQVVENGSFSRAAKQLSMANSAVSRVVKRLEEKLGVNLIINRTTRQLRLTEEGLQYFRRV
QKILQDMAAAEAEMLAVHEVPQGILRVDSAMPVHLHLLVPLAAKFNERYPHIQLSLVSSEGYINLIERKV

B45292

GATGAACCCCTCAATTGAAAATAGAGGACAATTATACATGTGGATGAGAGAAGTAATATCTATAACTCACC
CCAAATTATTCATAGCTGAAAATGTAAAAGGATTAACGAACCTTAAAGATGTAAAAGAAATTATTGAACA
TGATTTTGGTCAAGCTAGTGACGAAGGATACTTAATTGTACCAGCTTCAGTATTAAATGCTCAGTTTTAT
GGAGCTCCTCAATCACGTGAGCGTGTCAATTTTTTTTTTGGTTTTAA

5 **SEQ ID NO:74 polypeptide sequence of Orf37**

MKLISLFSGCGGMDIGFEGNFSCLKKSINEELHPEWISSTENEWTVSPTSFETIFANDIKPDAKAAWVS
YFLDQKANANEIYHLESIVDLVKKERETHNIFPKGIDILTGGFPCQDFSVAGKRLGFDShKNHhGKISNI
DEPSIENRGQLYMMWREVISITHPKLFIAENVKGLTNLKDVKIEIIEHDFGQASDEGYLIVPASVLNAQFY
GAPQSRERVIFWF

10 **SEQ ID NO:75 polynucleotide sequence comprising orfs1, 2, 3, 4, 5, 6, 7, 8 and non-coding flanking regions of these polynucleotide sequences.**

TATTGCAAACACTTCTCAGATGATTAAATAACATGGATACACGTTTGCCACACGGATTGCTGGTAACCTTT
GACAGTCGATGAAATAGGTGTGCTATGAGCCATTTATTTATTACCCAATGATGTGCAATGAAAAGATAGCGC
GTGCTATTATTCTTTGAAGATGATGCGATTGTATCGCACGAATTCGAAGCAATTGTAAAAGACAGTTTGAAGA
15 AAGTTTCAAAAATGTTGAAATTTTATTTTATGATCATGGTAAAGCAAAAAGTTATTGCTGGAAAAAACAC
TTGTCAAAAATTACCGTTTAGTTCACTATCGTAAACCCCTCTAAACGTCTAAACGTGCAATCATGTGTACAA
CAGCTTATTTAATTACTTTATCTGGCGCTCAAAAACCTCTACAAATAGCCTATCCTATCCGTATGCCGTGCTG
ACTACTTAACTGGTGCTTTACAATTAACCTGGACTAAAGGCTTATGGTGTTGAACCACCTTGTGTATTTAAAG
GCGCAATTTAGAAAATTGATGCAATGGAGCAACGCTAACAATGAAATTAATAAATAAATTACAAATGTTAAG
20 GTTGGGTCTAGGCAAAATATTTCTTGATAAAAAAAGCGGATTAAACAGAATAACAAATGTTCTAGAACAT
CCTCTTCTCCGCCAAGACGGAAAAATTTGGGGATTATGTGGTGAGCTCATTTGTATTCCGTGAGATAAAAA
ATTTAATCCCCACATTAAATTTGGTGTAATTTGTACCAAAACAAATGCTTATCTTTTAAACAAATCCATA
TATCGATCAACTTTACTATGTAAAAAGAAAAGTATTTTGGATTACATCAAATGTGGTCTAGCAATTCAAA
AGAACAATATGATTTAGTGATTGATCCGACGATTATGATTGCTAATCGCGATCTTTTACTTTTACGCTTAAT
25 CAATGCCAAGCATTATATTGGCTACCAAAAAGCCAATTATGGTTTATTTAATATTAATCTGGAGGGACAATT
TCACTTTTTCGGAACCTCTATAAACTCGCCTTAGAAAAAGTGAATATTACGGTACAAAGATATAAGCTATGACAT
CCCATTGATAAGCAAAGTGCGGTGCAAAATTTCTGAATTTTTCGAGAAAAACCACTAGAAAAGTATATTGC
TATTAATTTTATGGTGCTGCAAGAATCAAAAAGTAAACAATGACAACATCAAAAAATATTTAGATTATCT
CACGCAAGTCCGCGAGGAAAAAGCTGGTGCTATTAAGCTATCCTGAAGTAACAGAGAAATTAACACAAT
30 GTCAGCCGATTATCCGCATATTTTGTCCATCCAACAACCAAGATCTTTCATACCATTGAATGATTCCGCA
CTGTGATCAATTAATCTCTACAGACAGCTCTACTGTACATATTGCTTCAGGTTTTAATAAACCAATTATTGG
TATTTATAAAGAAGATCCTATTGCGTTTACACATTGGCAACCCAGAAAGTCGGGCAGAAACGCACATACTTTT
CTATAAAGAAAATATTAATGAGCTCTCACCTGAACAAATTGACCCCTGCATGGCTTGTCAAATAGTCTTATCT
CTTCTGACACTTGGGGCAATAGAACTATTTCTGTTGCCCTATCACTAACTTTCTATTTTTGTGCCACATGT
35 TGGACAAGGCTTATCCTTATTACCATAAACCAGCAATTTTGGACAAAATAGCCTGGACGCCATCCGGTTG
GAGAAAACCTTTTAGCGTCGTACCACCTTGTGGATTGCGTTAGACAGCACTTGTTTTTATTGTTCTACTAA
CTGCCCACATTGTGCTTGTAACTCCCTGCTGTTTTTGGCGATGTAGGTTACAAAGAAATAACGTTTC
ATTTCGCATAGATATTTCCCAACGCCAACGACGACAGCATTATCCATTAAAAAAGTTTTAAGTGCGGTCTGTTT
TTTACGACTTTTTTGGCCACAAGTAATCAGAATCAAATTCCTCAGACAGAGGCTCTGGGCCTAATTTTCAGAAA
40 AAGAGGAAATTCGTTCACTTCTCTGTCCATAACCACGCTCCAAAACGACGAGGATCGTTATAACGCACAAC
TTTTCCGTTATTCTACTACGATATCAAGATGATCATGTTTATCAATAAGATCCCCTTTCTCCACAACCTCTCA
TGACCCTGACATCCCTAAATGTCCAATCATAGCTGTTTCAAGTTGGATAATTAATACTTCGCACGGCG
ACTTAATGCGATGACTTTTTTGTGTGTAATTTGCGCTAATTCCTCGCTTACCATCCAGCGTAATTTTCGGTTG
45 GCGAACAACAATTTTTTCAATGATAGCCCCTTCAAGATAAGGGCTAATTCATTTTTTGTGGTTTCACTTC
AGGTAATTCTGGCATAGGTTATATATCCATAAATCTTATAATTGATAATATCCAACTATTTCATCAGCTATG
ATTGGCAGGCAAAAAGCCGCAATCGCGTAAATATTTTTGTCCGCAAGTCAAACAAAGCAAGGAGTCCACAAG
GCGTAATGCTTCCGCAGTAAAAGCTGCTAATGTATAGTTGCGCCCTCACATTATACTCATCAGGAATATCCAA
AACACAAATATCAGAATGCTGACGCAAGATTGATGATTACTCGCATAAACCAATGAAAAGATCAGCATAATT
CTGCTCAAAAAGCCACTCTGCGGTATTTCTGCTCTGTTGGAATAGTGATAGAATCCGGACCACCACTATTGC
50 CATTGCTTTTTTCTTTAATTCCGAGCCATAGCCCATATGCCGTTTTTCAATATTCGAAAATAATGCCAAAGT
ATAATCTCCACAAGGATCTGCCTTAGGTGTGATACCTTAAGCGTAAGTGGGGCGACATCAATAATGTCAA
CCAATTCATCATGGTGAGTAATCACCGATTTCTTTGCAATTAAACATAAACGATTGTAGCAAAAAGGCAC
AAGTTGAATATGAGGATATCGCGCTGTAAATGCCTAAGATGCGCATATTGGCAGAGGCAACAAATCCAC
TTTTTCCCTTGTCAATGCGTTGGCACAACAAACCCGCGGTCCAAATTCATTTTCGACTTGTAGGTGATA
55 CTGTTGGATTAAATGCTTGTGTCATAACGTAAAAGGCTGGCGTAAACTCCCTGCGGCTAAAATTCATGCG
ATATGTTTACTGTATGGTAAAGATGGGGACTAAAACCTGCTGTTCTCAATCATAGAATATTTAATCGGTAC
ATTATACGCTTGTTCAAATGAGATTCCGTTAAATTTGACTGGCTATTCCATATTTCCATTGTTGGTTAGG

B45292

AGCGCTGACCAATCTGGGCATGGGCTATGTATTTGAAGAACTGATTTCGTAAATTTAACTGAAGAAAATAACGA
AGAAGCTGGCGAACACTTTACCCACGCGAAGTGATCGAGCTGATGACGCATTTAGTCTTTGATCCGCTCAA
AGACCAAATTCGGCCATTATTACGATTTACGACCCAGCTTGCGGCAGCGGTGGCATGCTGACCGAGTCGCA
5 AAACCTTTATTGAGCAAAAATATCCGCTATCTGAATCACAAGGCGAGCGTTCATCTTTTTGTTTGGTAAAGA
AACCAATGATGAAACCTATGCCATTTGTAATCTGACATGATGATTAAAGGTGATAATCCCGAAAACATCAA
AGTCGGCTCAACCCTTGCTACAGATAGCTTCCAAGGTAATCACTTTGACTTTATGCTTTTCAACCCGCCATA
TGGCAAAAGCTGGAGCAAAGATCAAGCCTATATCAAAGACGGCAATGAGGTTATCGACAGTCGCTTTAAAGT
TACCTTACCAGATTACTGGGGCAATGTAGAAACCCCTTGATGCTACCCACGCTCCAGCGATGGACAGCTGCT
10 ATTCCTAATGGAAATGGTCAGCAAAATGAAATCGCCGAATGACAACAAAATCGGCAGCCGAGTGGCCTCCGT
GCATAACGGCTCAAGCCTGTTTACCGGCGATGCAGGTTTCAGGAGAAAGCAACATTCGTCGCCATATTATTGA
AAAAGATTTGCTCGAAGCCATCGTACAGCTGCCTAACAACTGTTTTATAACACAGGTATTACCACTTATAT
TTGGTTGCTGTCCAACAACAACTGAAGCACGCAAGGCAAGTTTCAGCTCATTGATGCCAGCCTCTTATT
CCGCAAAATTCGCTAAAAACCTTGGCGATAAAAACTGCGAATTTGTACCTGAACATATCGCCGAAATTACCCA
15 AAACTATCTTGATTTCACTGCCAAAGCGCGCGAAACCGACAGCCAAAATGAAGCAGTCGGCCTGGCTTCGCA
GATTTTTGACAATCAAGATTTTCGGCTATTACAAAGTCACCATCGAACGCCCGGATCGCCGTTCTGCCCAATT
TACCGCCGAAAATATCTCGCCTTTACGGTTTGACAAGGCTTTGTTTGAGCCGATGCAATATCTTTATCGGCA
ATATGGCGAACAAATTTACAACGCCGGATTTTTCAGCCAAACCGAGCAAGAAATTACCGCTTGGTGGCAAGC
GCAGGGCATAGCTTTAAACAACAAAAACAAGACCAAGCTGCTGGACGTCAAAACCTGGGAAAAAGCCGCCGC
ACTTTTTTCAGACGGCATCAACCTTGCTCGAACATTTTCGGCGAACAACAATTTGACGATTTCAACCAATTCAA
20 ACAAGCCGTGGAATGCCGTCTGAAAGCCGAAAAAATCCCCCTTTCTGCCACAGAGAAAAAGGCCGTTTTCAA
TGCCGTAAGTTGGTACGACGAAAATTCAGCCAAAGTGATTGCCAAAACACTCAAGCTCAAACCAAACGAATT
GGACGCCCTTTGCCAACGCTACCAATGCCAAGCCGACGAGCTGGCAGACTTTGGCTATTACGCCACCCGCCAA
AGCAGGCGAATATATCTATATGAAACGAGCAGCGACTTGGCGACAGCGAATCCATACCGCTCAAACAAAA
TATCCACGATATTTTCAAAGCCGAAGTGCAAGCGCACATCAGCGAAGCATGGCTGAATATGGAAAGCGTAAA
25 AATCGGCTATGAAATCAGCTTCAACAAATACTTCTACCGCCACAAACCATTACGCAGCCTTGCAGAAGTTGC
CCAAGATATTTTGGCGTTAGAAAAACAGGCTGACGGCTTGATTAGTGAAATTCTAGAGGCTTAATAAAAAAC
AACTATTAAGCAAGTTTTAATAGGCTTAAAGTAAGGAAATTCAAATATATAACACATTGAAAAATAATGA
ATTTTACCTTTTAAGCAAGATTTGGCATGAAATAAGCAAGGAATAATAATGACAGAACCCTTTCTAAATTT
AACGGCATTATCACAAAAAATTTATTTAGAGATGCAAGCTGACGCCGAAAAACCAATATTTTGAGCGCAAAGGACTAGGA
30 AAAAAAGACATCAAGCCAACTAAAAATAGCTGAAGAATTAGTTGGAATGCTCAATGCTGATGGCGGAGTTTTG
GCTTTTGGTGTGGCAGATAATGGCGAAATCCAAGACTTGAATAGCCTTGGCGATAAATTAGATGATTATCGG
AAATTGGTTTTTCGATTTTATTTGCACCGCCTTGTCGGATTGGACTGGAAGAAATTCGGTTGATGGAAATTA
GTTTTCTTATTCACGTAGAGCAAGATTTAGAGCGTATTTATTGTGCGAAAGACAATGAAATGTGTTCTTA
CGTGTAGCAGATAGTAATCGAGGCCCTCTCACCAGAGAACAATCAAAAATCTTGAATATGATAAAAAATATC
35 CGTCTATTGTAAGATGAAATAGTTCTCTGATTTTAAATGAAGAAGATTTAGATCAAGAATTATTAGAGCTATAT
AAAAAGAAAGTTAATTTTACCTCCGATAATATCTTAGATTTATTATACAAGCGAAATTTATTAAACAAAAAG
GAAGGTTGTTATCAGTTTAAAAAATCAGCCATTTTACTCTTTTCTACCATGCCGGAACGTTACATTCTTTCA
GCATCAGTCCGCTATGTTCTGTTATGAAGGTACAGTAGCGAAAGTCGGTACTGAGCATAATGTGATAAAAGAC
CAACGTTTTGAAAAATAATTTCCAAAGCTAATTGAGGAGCTGACCTATTTTAAAGAGCCTTTTAAAGGAT
40 TATTACTTTCTTGATGTCATCAGGGAAAAATTTATCAAAGTACCGGAATATCTTGAATATGATAAAAAATATC
GGTGTGTAAATGCGCTTTGTCATCGTTCTTACAATGTTTCAAGGTAATGTTATTTATATTAAACATTTTCGAC
GATCGTCTTGAAATTAGTAATAGTGGCCCTCTCCCTGCTCAAGTCACCATTTGAAATATTTAAACCGGAACGA
TTCGCTCGGAATCCACGTATAGCACGAGTTTTAGAGGATCTTGGGTATGTCGGTCAGCTTAATGAAGGCGTT
TCCCGTATTTATGAGTCAATGGAAAAATCATTATTGGCAAAGCCTGAATATAGAGAACAAAACAACATGTT
45 TATCTAACATTGCGCAACCGTGTTACCGCACATGAAAAACGGTATCTACAGCCACTATGCTGCAGATTTGAA
AAAGAAATGGACAAACTACAACGACACCCAAAAAGCCTTTGCTTTATCTATTTACAAATGGTACGGCGATA
TTGTACAGAAATTAGTTGACTATACAAAAATCAATCAGAATTCGATCCGAGCGTATTTAAATGCCTTTATTTCAG
CAAGGTATTATTGAAAGACAAAGTGTAACAGCGTGACCCCAATGCCAAATATGCTTTTTAGAAAAGATTAA
GCAAGGTTTATCGCTTGCTAAGCAAGGAAATGACAATGCTTAACTTGCTGAAAAATAATGATTTTTATCTT
50 TTAAGCAAGATTTGGCATGAAATAAGCAAGTTTTTTTTATAGTTAAACGGACAACAAATTGCATCAATAAGAG
CGGTCAATTTTTAAGGATTTTTTGCAAATGAGACGATACGAGCGTTACAAAGATTTCAGGTGTGGATTGGCTA
GGGGAGGTACCGAGCCATTGGGAGTTAAAACGCTTGAAACAAATTATTTGTTGAAAAAAAACATAAGCAAAGC
CTGTCTCTTAATTGTGGAGCCATTAGTTTTGGTAAAGTTATTGAAAAATCGGATGATAAAGTAACAGAGGCA
ACAAAACGTTTCATATCAAGAGGTGTTAAAAGGCGAGTTTTTAATAAATCCTTTAAACTTAAATATGACCTA
55 ATTAGTTTGAAGAAATTGCTTTATCAGAAATAGACGTTGTTGTAAGTGCCGGTTACATTGTTTTAAAGAAAAA
CAATAATTAATAAAAAATACTTTTCGTATTTATTACATAGATACGATGTTGCATATATGAAATTTATAGGT
TCAGGTGTAAGACAAACGATTAACATAGGGCAATTTTCAGACAGTATTTTGGTTATTTCCACTCTCTCCGAA
CAACAAAAAATCGCGCAATTCCTAGCATAAAAACCGCTAAAAATCGATCAGGCGGTGGATTTGGCGGAAAAG
CAGATTGCCCTGTTGAAAGAGCACAAAGCAGATCCTGATTCAAATGCCGTAACCCGAGGCTTAAACCTTGAT
60 GTGCCGTTAAAAGATTCCGGCGTGGAATGGATAGGGCAAGTGCCGGAGCATTGGGATGTGCAACGTTCAAAA
TTCATTTTCAAGAAAATAGAAAAGAAAGTGAATGAGGAAGACCAATTTGTTACTTGTTTTAGGGATGGGCAA
GTAACCTCTGAGAGCTAATCGAAGAACTGAAGGATTTACAAATGCGCTAAAAGAACACGGCTACCAAGGAATT

B45292

5
10
15
20
25
30
35
40
45
50
55
60

GCGAACGCTCCAGCTTCGATTTGCGCCTTCTCTAGAAAAATAAGAACGTATCTACTTCATCTAGCACCAATA
TTGCATTATCGGCTTTCGCTTGTTCAAAGGCTTGAGCAATATTTTGTCTGTCCCGCCCACATAAGGATTAA
GTAAATCTGAGCCTTGTCTTAGCAATAGCGGCATGTCCAACCTGTTCCGCAAGCCACGCTGCCCAAGCAGTTT
TTCCTGTTCCCGGCGGGCCATAGCAACAAATTCGCCCTTTTTTCGACCGTTTTTAACCTTCACTAATACGAT
GAATATTGTCGTTACAAGCCACATAATCCAAGTTGTAGTCGGCTTTGCCTAAAAACAAGCGGTTCAATTTTCG
GTTTATTTTTCGATTTTAAACGTTTGATTAAACATCATGAGCAAAGTCTCAGCAAAATTTGATGTATTGAGTT
CCTTTGCCACCCGAATTGTGCGGCTTAAAAATCGCCGGCGTTAATGACCGCACTTTAGCAAAATGCTGCACAT
AGGCCGGACTTAATTTTCCCTCAGTCAGTTGCGTAATCAGTGCTGACTTATTTTTCAACGGCAAATCTGGCA
TTTCTAAAAATAAAATCAAAGCGGCGTAAAAAGCAGGATCTATGCCCGAAACAGAGTTAGATAACCAAATCA
TCGGCACGTTATTGTTTTCCAATAACTGATTTGTCCACGCTTATTTTTTTTGTGCAACAGAACGCTCCATAA
ACGAGCCGTTAAACACATCTCAATTTTCATCAAAAATTAAGCGCCTGCTTGCCTTCAATAGCGTTTGAG
CAAGACGACTGTAGTTTCAAGCGTTGCTCTGCCTCCACAACATCTCCGTGAGAATCCATGTAAGTAATGTTAT
ACGCCGAAATCCCAACGCTGTGCAAGCAACCCGGCGAATTCTGTTTTACCAGTGCCAGGCACGCCATAAA
TTAAAGATTACGCGCTTTTCGATGATGTTTTAGTGCTTGTGTCAAATAAGTCAACATCATCTCTTTCATGC
CGGCAATATGGTCAAATCATCCAGTTGCAGACTTGGCACTTGAGCGACTTCCGTACAAGATTTTAAATAGGA
CGTTTTTCGTTAATGGTTGTGTCAAAATTCATCAAAATCTAAGGTTTCGCCCCAATCTAAATAATCATGCA
CACTATCGGGGCGATAATCGCGATCAATCAGGCCATAAGCATCGAGTTTACTGCCTTTCTTTAAGGCAGATA
GAATCTGATTTTTCGGCTGTTAAGTAAATCCGCCATGATCGCAGCCGTTCTTTGTAAATCCGATTTTCGGCA
AGTAGCCAAATCTCGCATAGCTCCTTCACTACGTAAATGCATGGCAAAGCGGAGAAGTTCTCTGTTCAA
CGGGATTAGTTGCAAAAATTCGCCAACGTTGCCAAATTTTCATACGCTGTTTCCATAACTCAGGTAAAA
GTGCGGTGGATTTTTGGAGTTTATATACCGCTCTTTTAAAAGCCGACGAGCAACCGTGCCTAAATTTTTAT
CATTCTCTAATTCTTCAGGCAGCCCAAATGCACCTGGCAATTTTCATCACTTCGCCAGCTAGTCTCCGAAACA
CTTCGGAAAAATCTTTATGCTCAAATAAAATTTAAGCATCATATTTTCAGTATAAGAACACACTGCTCGGTG
GGTTAAATTTATATTACAGCATAAAAAATCTCTTACTGGGTTGGTAAGGAGTATTTTAGTGAGTAGTG
GACAAAGGTTCGTTAAGGATAGTTTTAAGAAGCTTTGTTAATCAACCATCAACTAAACCAGCACTAATT
ACAAGCTCTGCCATTTTCGGCCATTTACAAGCTTAATTCCTTTAGCTTGATATTGTTTTTCATCTATGTTT
TCTAAACCAGAATGATATACGTAATACATTTCTGAATAACCATAGTTTTTATATTCACTTTCAAAGTTCGAA
ACATATTTCGTCTAATTGTTTAAATATCCGTATCTGACTTAATTTGCACAAATACTCTCTCTGCGTTGAAGAC
GAATACAAATCAAGATCTATTCCTTTCTCGGTTTACCTAAAAACAGAGTATCGTTGCCATCCTAATTTAGAA
AAAACAAGATCCGTTAAAGTTCAAAGTCACTCCACCATAAACCTTTAATTAATTTTCAACTGATTTAATT
AATGTTTCATAACGCTCTTTTCGCTTCTGTAATTTCTCAATAACTTCACCATTATACGACGTATTAAATAG
TCCTCCATCTCAACACCACAAATCGTCCCTCTATAGGCTTGGACCTTTGTTACTCTACCATCAAGATTATCG
ACTAAAAGCTCTTTACCGTTAGCATCAACGCAAGACCAATTCCTATGTTACTAATAACTTTTCTGTTCTA
GAACCATCGCTTTTCCCAACAACTCTTTACTGCAAAAAGCCCAATATAATTTACGTCCAAGAAGGTGATC
CAAAGTGATCTTCCCAAGTTGATAAAAATCTGAATTTGTCTCAAGTGATTTGAACAGTTTCTGTATGG
TCACTCCAATAAGTTTTTACAATATTCAATACAACTATCCCATTGATTATTCAAACATTCTTTGTGAATCTCT
GATGTAGATTTCATAGCCAAGACGAATCGTATTTTTTGTACTTGCTGTACTATTTTTATCAATACAATCTTTT
TCCCAACATCCTTTTATGCCTAATTTAATAAAACGAATATTAGTAGGTTCAATTTTTTCAAACATAGTTTTT
CCTTATTTCTAGTTAAATTCACCGAATTATAGATAATTGAGCAAAAAAACAATTTAAACATATTTTTT
ACTCAATAATAGAAATGACAACAACTACCGACAAATCATCCGAAAACGATTGCTTCTCAATCTTGC
AAACCGTAAGCGCATATTTATCATCGGGATTTTTCGCCAAATTTTTTCGCGCATTTTCATCTGAAAGCCCGT
CGGTCAAGCCGTCAGAACAAAGTAATAAACTTTCCCTTGTCTGAATTTCAATTTCTTGATAAAAAATTTTAT
CTTGAAATTCGGAATAATCGGCGACTAAACAAGAAGAACGCCGCCATAAATCGTGGCAAAATCTTCTCTT
TTTTATCGGGGAAATCAGTCAATAATTCAGAAAGAATAGAAATGATCTTGGGTGATTTGTTGCCATTTTCTT
GGGCATCAATTAAATAAGCACGACTATCGCTACGCTGAGAATTTTCGCTTTACGGGTATTGATCAATTT
CGGCAGCCACAAATGTGGTCGCCGAACCAAAATATCCTCAGCTAATTCGTGCTGATAAACTGGATGTAAAT
CGTAGATCGTTTGACGGTTTATACTTTCCATTTGGCTTAATAATTGCATAGCCAATTTGCTCGCTTTTTCAG
GTCGGTTGCTATTAGAAATACCATCTGCCACGCCCAATAAAGTGCGGTGCGTTTTCAAGGCGTTTTTCAG
CCGTTTTGAGTTTTATATTGAAACACCGCTCGCCATTAAAAAGGGCATCTTGGTTGCGTCGCTTGTGCTGC
CAATTTTGTGTTGGCAAAGGGTAATTTTCGCAAAAATTTTTTCATTTATTCAACCGCTTGTGAGAAGGATTTAA
AGGCGATCAATCGCTTTTAGTGATCTAACGCTTTTCATTTCTTAGACTTAAAAAAGTGCATTTTCGGGCACG
CCCTGCATCTTGTGGGGTAATACGGGATAACCCCCCTTTTTTTTTGCTTTTTCGCGGTACGTTTCAGAAATC
GACGCACAGTGAATGGCTTTTCTGTTCCAGTTTCGATAACGACGAGATTTTGCACCTCTTTTAACCACGA
TTCTAACCGCACTTTTTTAAATCCTGATATTGACTTGACATAACTCCAATCATTAAACATTAGTACATTTTG
ACGAGCAAAGCCCCACAATAAGGCAAAATGTGGTTTTTCACTGGTTAAACATAAGTTTTTCATTATCCACGAC
AGGTTGAAAACCTTGATGCAGACCAACTTAATCCTCGACAATTATTGACACATTGAAGACGCTCCAAGTACC
ATGTACTTCAATAAATGGCTATCATTAACACCCAGCTTTTGAATAATGCCCATCAACATCTGGTAAAC
AAATATCCATGAGGTTTATCTCCCGCCAGCATTTTAAATCTGATACCCTTCGTGAGGAAGAGTATTTTCG
GTATTGAACTAATCGATGCCCATAAACCAATAGGCTAGTTCTGATTATGCTTATAAGCTAGTGGCGTTGC
GATCTCTTCAAAGATATATTATGTTCTTTAAACATAGGATAAGCATTCCAAAATCCGCCAACGCTGCGGAA
ATCGGGAAGCCAGAATCCACGCTCATACCCGACACGCTGTAATTAATAATGCCATCCGCTTTGCGGATAAG
TTCCACTGCATAATTCAAATCATTTTTTCATAATACTTTTCTCTGCCATTTTTTCATGATGAAATAATACCC

B45292

ATCTGTTGTAAATAATCTTGGGGTAGGAAAGCGCCAACCAAATTCGACGACCTTGATAGAAAAGCATCCC
TTGTTTGTCTTTCATTATAAGTTTCTCCTTTTGGAGATTGCCCCATAACGACATCATAACAATCGCCAATCGT
AGATAATCCCACCCCTTCGGCACTTCAACCCCATCAACCTCCACCATCTCACACGGAAACGCTTTGGCGGT
TTCGGCTAGTTTCGGCGTAGCGGTGAGGCTGTGTTTGTGAAAGTGCGGTGAGTTCTTCGGGTGTTTTTCCGCT
GATTGCCCTGCATGGCGGCAAGTTCTGCTTGTTCAGGCTAAGACCGTCTGAAAGGGCTTGGATTTTGGCAG
CACGGGATCGAAATCGACAAACAGCTTTTAAACAGGGCTTGGGCGATTGTTCTAAGGTTTGGTTGATTTG
AGTGTGAGTTCTATTTTTTGTATCTAAAGTATTTAGAAATATATCCAATTTCTTTTGTCTTATTTATATCTAG
AAGTAATAATTTAACTTTACTTAAAGCTGATACATATAGATTTTTTTTGGACACTACCTTCAGCTAAAGTTTC
AATTTGTTCTTTTGTCTATTTAATAAGGTATAAAAAATAAATAATGGGTTACAAATATTTTCATTAACTTTGAT
ATTAATTACAGCTCTATTTCCACACATGTAATCTTTTAAGATTCCAATTCGTCCAATAGTTCCTGATTTGCT
AATTGCTAAACTATCTGGTTCAAATAACAGCACTCTTCTTTGCACTTAAAAATCCTTTTCTGTTTAAAGT
TTGAGAGGTTTTATATACAAACCATGTTTAAATCTGTTGCTCTCAACCATTTAATTGTTCTCTCAAAGTA
GCTTGGTTTCAATTTTTTGTATGATTAAACATAACCTTCTTGAAATGAAGCGCAATCAGCTAAAGTTATAACCTT
CCAATCATTAT

SEQ ID NO:79 polynucleotide sequence comprising orf25 and non-coding flanking regions of these polynucleotide sequences.

CACGCTAGTGCCGCTCAATCCGACGCGACTGCGTCGCAATCGGTTAATCATAAGTGAGTGGCGTTGCCACT
CGTGTGGGAGAACACAGCCCCAGCGGGGCTGAATTATGCGTAACCATGTACGGCTTTGCCGTGCATGGGAA
AAAATAAGCGGTGAAATCTTGCAAATTTTTTGCAAATCTTACCGCTTGTTCTTTTGAAAAAGCATTAAAA
CTCATCTAAATCATCTTCATGATTCATTGATTTTTTATGTCCGTTATCCATTCTTATATTTAATTGCAAGTTC
CATATAATCTTTATTTCTAAGTTCTTCATCTTCAGCTATTTTTTCAATTAACTATTTACTTTATCCTCATC
TCCACAAATTTTAATTAAGGCATCCCAAAGTAGAATTTTCTCTCTATGTATGTTAGGATCATCCCCCTTTTG
AGATTTACGTTCGTATATTGAAGATTTAAGTAATGATAAAAAATACTTCAGGGGAACCTTAATATATCATCAGA
AAAAGGGTTTCCAATTGAAACAAAGAAATAGATTATGTGACAAATATAGGTTCCGTGCAAGTTCTTTAAT
TGTTGACAGAGCAATATTTTAAAAATATTTTCTCCAAGTCTTTTGCATCCGATTCCAGATACTAATTGATG
ACCTCGGCCCTCTCGATATCCAATAATCTTACAATTTGATATTGCCCATATAGATCGCTAGAATTTAATAG
TTGAGTAATAACTTCTTTTTTATCCTTCTCAGGAAGTCTTCTAAGTAATCTATAAACTAAGCGACTCCAAAC
CATATCCGCCCCAAAGTCAAAGAATCCTAATTTCTTTTTCAGGCACCTCTTGGTAAATTTCTATATAATGTTGG
TATAGTTGCTAGAGCTATTTCTTTAGTAAAGTCTTTTTTCATAGTCAATTAAATGTTAACTACATTTTCTAG
AGAATCGTCAGGAACAGCTGATAAAGCGATCTTGAAATCTTCTCTGACTGCATTGCAAGCCAAACTTTTTG
TGATAATTTAACATTTTATGAATCAGGACTCAATACTTGTTCAAAATATAAATCAAAGAATGCCGAATAAGC
AATCCTTCTATTTTTTAGGAATTCATTATTTGAATTTATATTATCAATATCAAATAAACTTCTAGAAAAGA
CTCATACATTTTCAATTATCTTGAATAAAATCACTTAACTTAACTTTTTCTTTTGTCTTATCTGATCGTGCCAA
GAGATAATCTTTAAGTTCAAAAATTTCTTTAAATTTATCTGGAAAGAAAATCTTATCGCTTCAATAGTGAG
TAAATCAACCACATCAATTTCTTTACCTAATTTGTTTAAAGATATTCGATAGAGAAGATGTGTAACGCTTAAT
ATCTCGAATATTTTTTATTGTTGGCTTAATGATATTTCCAATATGCAATTAGACCAACGCGCTTATGAGTA
AACATCCCTTAAAAATCTTATCTAAAGATGAAAAATAAATTTCTTGTAAATAGTTTTTTAGGTACCTGTGGTAT
ATCGAATGGAATCTGAATTATCTTCTTAAATAATCCTGGCCATCAATGGTATTATCATTTAATGGTTTAAAT
TACTCTATTTTTTATCAAATGATAAAACATAAACAATATTAGGAAAGTTTCTGTAACTCTGACCAATTTTAG
AATTGATTGTAATTCATCAGATGATAAACGCTTATATCATCTAAAATTACAGTAATAGGTTTACTTATTTTC
CTTTAGAACTTTAATTAATTTATCACGTTGATTTTTCAAACGTGTTTTTTCTTTCTTTTGTGAAAAAAA
ACTTAAACAGCCACCCAAAGACACTAAAATAAATTTCTTCAAAATGGAATAGGTTTTAAATTAGATAACAACCTC
TCCAAAACACTACTCAAACTATCAATTAGCTCATTATCATCTCATAATCTCTTAACTGAGCAGAGATTTTCAGT
AAAAAATAAAGCAACTAAGTTATGAGCATCACTAAACATCCAAGGATTAAATCAAGTACAAAAGAATTTTTT
TTCTAATTTCTGGTCGATTAATTTATATAGGATGTTTTACCATTTCCCCATTCTCCACATAATCCCACAAC
CAAACCTTCTTTATAGTCAAAATGAAAAATGTGTTTAGCAAATGCTTCTGCACTACTAGCTCTACCTAATAA
ATCATTGCTAGAATCTTTTATTGGATTATCGCTTATTAATTCATATATTTTCTTTAGTAAATGCTCATAT
CTTTTATGTGTAACC

SEQ ID NO:80 polynucleotide sequence comprising orfs26, 27 and non-coding flanking regions of these polynucleotide sequences.

TTATTGAATTTCCCTGGCAGAGAATAATATGACAAAAGTTTAGACAAAATTGCAAAACAATTAAGAGATTCT
GATAAAAAGGTTAATCTAATTTACGCCTTTAATGGAAGTGGAAGAACCCGTTTATCAAAGTCTTTAAGAAT
CTTATTGCACCTAAAGAAAATCATGCAATGAAGAAGATCTAACACGAAGAAAATTTCTTTATTTCAATGCC
TTTACCGAAGATTTATTTCTATTGGGATAATGATCTACTAATGACACAGAACCAAAATTAAGATTCAACCA
AATCTTTTATTTCGCTGGTTGATTAGAGATCAAGGGGATGAAGGTAAAGTAATTGGAAAATTTTCATCATTAT
TGTGATGAAAAACTTATGCCTAAATTTGATATAGAAAATAATCAAATTACATTTCAGTTTTGCACGTGGAGAT
GATACGCCTGAAGAAAATATAAACTATCGAAGGGGGAAGAAAGTAATTTTATTTGGAGTATTTTTTCATACG
TTAATTGAACAAGTTGTTGCAGAATTAATATCTCAGAGCCTAGTGAACGCACTACTAATGAATTTGATGAA
CTTAAATATATCTTTATTGATGATCCAGTAAGTTCAATTGGATGAAAATCATCTTATTCAATTAGCTGTTGAT

B45292

AACTCATCTCTTAAACTAGTGTATTACTCCATGTTTAGTCACTAGCCATAGTGCGAATTTATCATATTTA
TTTCTAGGATTTTCTAAGATCGTTTCAGGGAAGAAAGCATATGCTTGAGCAATTAACATATCTCGCTCATAT
TTTTCTATTACCTTCCAGTGTCTAAGTTTGGGAGGGGCAATATCATCTAAACATTTGTTTGAAGTCCACCAT
5 AAACCTTTACCTTCTATTAAATAGAGATTTATAGTATTTAGCTACAGGAGTTATTGGATTATCTAATTCTCGG
AGAGAATCATATGTGGTCTCCATTTTTCAAATATGCTCTCCCCCTTTCTAATAACATATCTATTTTATAT
CTAGGGTAATGGGTACAGCTATATCACATAAACACGACTCATATGGTTTGGATAGGAATTTAATATTTCCA
CCTAATTTACCAATGTAAAGAAAATTTTTCTATATTTGGAATTCGTGTAGACTCAAGAATAGAACTGCCA
ATTGAAGTCCATTTATCTCCTTGTGTACTTTTTACTTCAATACCATAATATTGACTAGCTACAATGTCTGGA
10 AAATGTTTCCCTGATACTAACTAATAGTGTCTTCGAAAGGAGTATTTGAGCACAATAACAAATAGCCTCA
TATACATCTTTTCTAATCAATACCCTACGTTTCTTATAGTATGCAACCCTATTTTCTGCATCATGATTA
AGAAAATATCGACTCTATTCATTAATGACGTGAATTCATGTAAAGGTGGATACTTATTTTATAGAGAAAATC
ATAAATAAATCCTATTTTAAATAAAGATTCCATTTTTTTCTATATTTTAGCAATATTATAAGCTAATATA
CATGGTACCGCATTTGCCAATAATTTTATAGGCATTACTGGCTGAAACAGAAACGTTTTCTGCTGTTTTAGGT
15 AAAATAAATTGGTATCTATCAGGAAACGTTTGTAACTAGCACATTCCTTTATAGTAAGACGACGTTTCGAGC
ATGCCCTTAGATAATTCGTTAATATATTTCCCTTCATGCTCTATGCTTAGCCTACGATTTTCAATATTACCA
TGATGTTTCAAGATCGGAATGTTGGGGCCCAACAGAAATTAAGTTTTAAATTTTCAACCCCTGGCCCTTGG
ACCAATGGGTTTTCCCCATAAATTATTTGGGGCTTTGGGGAAATAATTTTTTGGTTTTGAAAAAAGGGGGT
TCTTTTGGTTTATAAAAAATTTGGGGTTTTCTTTTGGGAGGAATTTATATTAAAAAGGGCCCTTTGGGGGCG
GCCATTGGGTAAACCAACCCAGACTTTTTC

SEQ ID NO:83 polynucleotide sequence comprising orf33 and non-coding flanking regions of these polynucleotide sequences.

ATGTTAAGGCTTGAGGCAAAGAATGGGCTCAAGCCTTTTGATTTTCATCAAAATATAAAAAATTAAGGAGATTA
TATGAGTGTACTCAGTTACGCACAAAAAATCGGTCAAGCCTTAATGGTGCCTGTGGCAGCCTTACCTGCTGC
25 TGCATTATTAATGGGTATTGGCTATTGGATCGACCCAGATGGTTGGGGTGCAATAGTCAATTAGCCGCATT
ATTAATTAATCTGGCGCAGCAATTATTGACAACATGGGCTTACTCTTCGCTGTGGGCGTCTGCTTTTGGGCT
TGCAAAAGATAAACACGGTTCCGCGGCACCTTTCAGGCTTGTGGTCTTACGTAGTAACCACCTACTTTTC
CCCTGCTGGTGTAGCACAAATTACAACACATTGATATTAGTGAAGTGCCTGCCGCATTCAAAAAATCAATAA
CCAATTTATTGGGATTTTAATTGGTGTGATTTCAAGCTGAACCTTTACAACCGTTTCTATCAAGTTGAATTACC
30 AAAGGCACTTTTCGTTCTTTAGCGGAAAAACGCCCTCGTCCCAATTTTGGTTTCTTTCGTGATGATCGCCGTATC
ATTTGCCTTACTCTATTTTGGCCTCATATTTTAAACGCTCTCGTTTCATTGGTGAATCCATCAAGATTTT
AGGTGCAGTAGGTGCGGGGATCTACGGTTTCTTCAACCGCTTATTAATTCCTGTAGGCTTACACCATGCCTT
AAACTCTGTATTCTGGTTTGATGTAGCGGGTATCAACGATATTCCAACTTCTTGGGCGGCGCTAAATCCAT
TGCCGAAGGCACTGCAACCGTGGGGCTAACTGGTATGTATCAAGCTGGTTTCTTCCCTGTCATGATGTTTGG
35 TTTACCAGGTGCTGCTCTTGCAATTTTACTGCGCAAAACCAAACCAAAAGTACAAGTGGCCTCAATTAT
GCTTGGCGGGTGCCTTAGCCTCTTCTTTACAGGGATCACTGAACCGCTTGAATTCTCATTATGTTCTGTTGC
ACCTGTACTTTTATGTATGTCATGCATTATTAACAGGTATCTCTGTATTTCATTGCAGCTTACAATGCAGTGGAT
TGCGAGATTCCGATTTAGTGCAGGTTTAGTGGATATGGTACTTTCTAGCCGTAACCCACTTGCCGTTAGCTG
GTATATGTTACTTGTACAAGGTATTGTATTCTTTGCTATCTATTATTTTGTGTTCCGTTTTGCAATTAATGC
40 CTTTAATCTCAAACGCTAGGACGTGAAGATAAAGCGGAAACAGCTGCAGCCCCAACTCAAAGCGACCAATC
TCGCGAAGAAAGAGCGGTGAAATTTATTGCTGCTTTAGGTGGTTTCAAGAACTTCAAACCTGTGGATGCTTG
TATCACTCGTTTACGCTTAACTTTAGTTGATCATACAATATTAACGAAGATCACTTAAAGCGCTTGGTTT
AAAAGGTAATGTAAATTAGGCAATGATGGATTACAAGTCATTTTAGGGCTGAAGCTGAACCTGTGGCAGA
TGCG

SEQ ID NO:84 polynucleotide sequence comprising orf34 and non-coding flanking regions of these polynucleotide sequences.

GGGATTTTCATTATGCTGTTTTACTTTTATACTTTAAAGGTGCAAAAAATAAAAAACTCTTTTTCGCTAAACGG
AATAATAAAATGAAAACAACCTTCTGAAGAATTAACGGTATTTGTGCAAGTAGTCGAAAATGGCAGTTTCAGC
CGTGCGACCAAGCAGCTATCAATGGCAAATTCGCGGTAAAGTCGTGTGGTGAAAAGGCTAGAAGAAAAATG
50 GGTGTGAACCTAATCAACCGCACTACTAGACAGCTTAGACTAACAGAAGAAGGCTTACAATATTTTCGTCGC
GTACAGAAAATTCGCAAGATATGGCTGCAGCTGAAGCTGAAATGTTGGCAGTGCACGAAGTCCCAAGGC
ATACTACGCGTAGATTACGCCATGCCGATGGTGTACATCTGCTAGTGCCACTGGCAGCAAAATTAACGGAA
CGCTATCCGCATATCCAACCTTTCGTTAGTTTCTTCTGAAGGCTATATCAATCTGATAGAACGCAAGTCGAT
ATTGCCTTACGAGCTGGAGAATTGGATGATTCTGGGCTGCGTGCTCGTCATCTATTTGATAGCCACTTCCGC
GTAATCGCCAGTCCAGACTACTTGGCAAAACACGGCACGCCACAATCAACTGAAGCTTGTGCCAACCATCAA
55 TGTTTAGGCTTCACTGAGCCAGTCTACTAAATACATGGGAAGTTTTAGATGCTCAAGGAAATCCCTATATAA
ATCTACCGTACTTTACCGCCAGCAGCGGTGAAATTTTACGGTCATTGTGTCTTTTACGGCTGTGGTATTGCT
TGCTTATCAGATTTTTTGGTAGACAATGACATCGCTGAAGGAAATTAATTCCTTACTTACTGAACAAACC
GCCAATAAAACGCTCCCCCTTCAATGCTGTTTACTACAGCGATAAAGCAGTCAACCTTCGCCTACGTGTGTTT
TTAGACTTTTTTAGTAGAAGAGCTAAGGGGATAATTAAATTCATAGCATTGAATTTTAAAGTCAATTTGCAA

CLAIMS:

1. An isolated polypeptide comprising an amino acid sequence which has at least 85% identity to an amino acid sequence selected from the group consisting of SEQ Group 2 ,
5 over the entire length of said sequence from SEQ Group 2 .
2. An isolated polypeptide as claimed in claim 1 in which the amino acid sequence has at least 95% identity to an amino acid sequence selected from the group consisting of SEQ Group 2, over the entire length of said sequence from SEQ Group 2 .
10
3. The polypeptide as claimed in claim 1 comprising an amino acid sequence selected from the group consisting of SEQ Group 2.
4. An isolated polypeptide of SEQ Group 2 .
15
5. An immunogenic fragment of the polypeptide as claimed in any one of claims 1 to 4 in which the immunogenic activity of said immunogenic fragment is substantially the same as the polypeptide of SEQ Group 2 .
- 20 6. A polypeptide as claimed in any of claims 1 to 5 wherein said polypeptide is part of a larger fusion protein.
7. An isolated polynucleotide encoding a polypeptide as claimed in any of claims 1 to 6.
- 25 8. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 85% identity to an amino acid sequence selected from SEQ Group 2 over the entire length of said sequence from SEQ Group 2; or a nucleotide sequence complementary to said isolated polynucleotide.
- 30 9. An isolated polynucleotide comprising a nucleotide sequence that has at least 85% identity to a nucleotide sequence encoding a polypeptide selected from SEQ Group 2 over

17. A process for producing a polypeptide of claims 1 to 6 comprising culturing a host cell of claim 16 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

18. A process for expressing a polynucleotide of any one of claims 7 – 14 comprising transforming a host cell with the expression vector comprising at least one of said polynucleotides and culturing said host cell under conditions sufficient for expression of any one of said polynucleotides.

19. A vaccine composition comprising an effective amount of the polypeptide of any one of claims 1 to 6 and a pharmaceutically acceptable carrier.

20. A vaccine composition comprising an effective amount of the polynucleotide of any one of claims 7 to 14 and a pharmaceutically effective carrier.

21. The vaccine composition according to either one of claims 19 or 20 wherein said composition comprises at least one other non typeable *H. influenzae* antigen.

22. An antibody immunospecific for the polypeptide or immunological fragment as claimed in any one of claims 1 to 6.

23. A method of diagnosing a non typeable *H. influenzae* infection, comprising identifying a polypeptide as claimed in any one of claims 1 - 6, or an antibody that is immunospecific for said polypeptide, present within a biological sample from an animal suspected of having such an infection.

24. Use of a composition comprising an immunologically effective amount of a polypeptide as claimed in any one of claims 1 – 6 in the preparation of a medicament for use in generating an immune response in an animal.

ABSTRACT OF THE DISCLOSURE

5 The invention provides BASB231 polypeptides and polynucleotides encoding BASB231 polypeptides and methods for producing such polypeptides by recombinant techniques. Also provided are diagnostic, prophylactic and therapeutic uses.